



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY

INSTITUTE OF INFORMATICS

SHLUKOVÁ ANALÝZA 100 ČESKÝCH FIREM NA ZÁKLADĚ ÚČETNÍCH VÝKAZŮ

CLUSTER ANALYSIS CZECH 100 COMPANIES ON THE BASIS OF FINANCIAL STATEMENTS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Tomáš Langer

VEDOUCÍ PRÁCE

SUPERVISOR

Mgr. Veronika Novotná, Ph.D.

BRNO 2017

Zadání diplomové práce

Ústav: Ústav informatiky
Student: **Bc. Tomáš Langer**
Studijní program: Systémové inženýrství a informatika
Studijní obor: Informační management
Vedoucí práce: **Mgr. Veronika Novotná, Ph.D.**
Akademický rok: 2016/17

Ředitel ústavu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává diplomovou práci s názvem:

Shluková analýza 100 českých firem na základě účetních výkazů

Charakteristika problematiky úkolu:

Úvod
Cíle práce, metody a postupy zpracování
Teoretická východiska práce
Analýza problému
Závěr
Seznam použité literatury

Cíle, kterých má být dosaženo:

Provedení finanční analýzy veřejně přístupných účetních výkazů vybraných 100 českých firem (na základě dat za rok 2015) pomocí vytvořené aplikace a potvrdit nebo vyvrátit existenci vztahu mezi ukazateli výkonosti firem a jejich klasifikací z pohledu provozované ekonomické činnosti a velikosti.

Základní literární prameny:

HLAVENKA, J. Vytváříme WWW stránky. 6. aktualiz. vyd. Praha: Computer Press, 2002. 369 s. ISBN 80-722-6748-5.

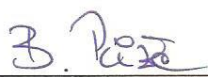
KNÁPKOVÁ, A., D. PAVELKOVÁ a K. ŠTEKER. Finanční analýza: komplexní průvodce s příklady. 2., rozš. vyd. Praha: Grada, 2013. 240 s. ISBN 978-80-247-4456-8.

LACKO, L. Databáze: datové sklady, OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle. 1. vyd. Brno: Computer Press, Brno, 2003. 486 s. ISBN 80-722-6969-0.

ŘEZANKOVÁ, H., D. HÚSEK a V. SNÁŠEL. Shluková analýza dat. 2., rozš. vyd. Praha: Professional Publishing, 2009. 218 s. ISBN 978-80-86946-81-8.

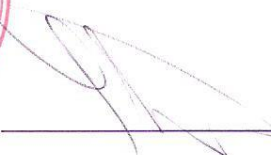
Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2016/17.

V Brně, dne 28. 2. 2017



doc. RNDr. Bedřich Půža, CSc.
ředitel





doc. Ing. et Ing. Stanislav Škapa, Ph.D.
děkan

Abstrakt

Diplomová práce nazvaná Shluková analýza 100 českých firem na základě účetních výkazů se zabývá testováním dvou stanovených hypotéz za využití vícerozměrné statistické metody – shlukové analýzy. Jako vstupní podkladová data k aplikaci statistických metod slouží veřejně dostupné finanční výkazy vybraných společností za období 2014 a 2015. Tato data jsou digitalizována a podrobena metodám finanční analýzy.

Abstract

The diploma thesis called Cluster Analysis Czech 100 Companies on the Basis of Financial Statements deals with the testing of two hypotheses using a multidimensional statistical method - cluster analysis. The input data for the application of statistical methods are financial statements of selected companies for the years 2014 and 2015 which are publicly available. These data are digitized and subjected to methods of financial analysis.

Klíčová slova

Shluková analýza, finanční analýza, R, finanční výkazy, rozvaha, výkaz zisku a ztrát, digitalizace dat

Key words

Cluster analysis, financial analysis, R, financial statements, balance sheet, profit and loss account, data digitization

Bibliografická citace

LANGER, T. *Shluková analýza 100 českých firem na základě účetních výkazů*. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, 2017. 75 s. Vedoucí diplomové práce Mgr. Veronika Novotná, Ph.D.

Čestné prohlášení

Prohlašuji, že předložená diplomová práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu Zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 26. května 2017

.....

Langer Tomáš

Poděkování

Rád bych touto cestou poděkoval vedoucí mé práce paní Mgr. Veronice Novotné, Ph.D., za ochotu a cenné rady, které mi pomohly při zpracovávání této práce.

OBSAH

Úvod	10
Cíle práce, metody a postupy zpracování.....	11
1. Teoretická východiska práce	13
1.1 Účetní výkazy.....	13
1.1.1 Právní minimum	13
1.1.2 Povinnost zveřejnění účetní závěrky	14
1.1.3 Rozvaha	15
1.1.4 Výkaz zisku a ztráty	16
1.1.5 Přehled o peněžních tocích	16
1.1.6 CZ-NACE.....	16
1.1.7 Klasifikace firem dle počtu zaměstnanců	17
1.2 Finanční analýza.....	18
1.2.1 Ukazatele finanční analýzy.....	19
1.2.2 V podnikatelské praxi nejčastěji užívané finanční ukazatele.....	20
1.3 Digitalizace dat a ETL	20
1.3.1 Optické rozpoznávání znaků (OCR).....	20
1.3.2 ETL proces	23
1.3.3 Technologické minimum	24
1.4 Shluková analýza	27
1.4.1 Principy shlukové analýzy	27
1.4.2 Metody rozkladu.....	30
1.4.3 R	32
2. Analytická část.....	33
2.1 Analyzované společnosti.....	33
2.2 Zdroje dat	42
2.2.1 Kvalita a dostupnost podkladů.....	42
2.2.2 Do analýzy nezahrnuté společnosti.....	44
2.3 Proces zpracování dat.....	44
2.3.1 Digitalizace podkladů	44
2.3.2 Datový editor	45
2.3.3 Provedení analýzy nad daty	47
2.4 Výstup analýzy dat	50
3. Návrhová část.....	52
3.1 Vstupní datová matice	52

3.2	Hierarchické shlukování.....	53
3.2.1	Na základě činnosti subjektu	55
3.2.2	Na základě velikosti subjektu	58
3.2.3	Analýza výřezu dat	60
3.3	Ploché shlukování	61
3.3.1	Stanovení počtu shluků.....	61
3.3.2	Vizualizace dat	62
3.4	Interpretace výsledků	63
	Finanční náročnost řešení	65
	Potenciál praktického užití	68
	Závěr	69
	Seznam použitých zdrojů	71
	Seznam tabulek	74
	Seznam příloh.....	75

ÚVOD

Autor se v diplomové práci zabývá klasifikací vybraných obchodních korporací pomocí hierarchických a plochých metod shlukové analýzy ve snaze nalézt určité vazby mezi těmito obchodními korporacemi na základě ukazatelů ekonomické výkonnosti.

Shluková analýza patří mezi metody třídění a klasifikace objektů, umožňuje nám tak nalézat mezi objekty neočekávané vazby a ty dále rozkládat, stejně jako umožňuje nalézt charakteristického představitele určité skupiny a získat tak povědomí o chování libovolného zástupce dané skupiny. Můžeme však také hlouběji členit již vytvořené celky, čímž nám začnou vznikat pomyslné stromové struktury. V praxi a běžném životě lze nalézt nepřeberné množství různých třídění a klasifikací. Díky výpočetní technice a nárůstu výpočetního výkonu se celý obor klasifikace a shlukování dostává do popředí. Pravděpodobně největší pozornost je v současnosti věnována shlukování rozličných dokumentů, například za účelem nabízení co možná nejrelevantnějších výsledků při vyhledávání na Internetu.

V předložené práci autor aplikuje poznatky získané z oblastí shlukové analýzy, finanční analýzy, databází a programování.

CÍLE PRÁCE, METODY A POSTUPY ZPRACOVÁNÍ

Hlavním cílem této diplomové práce je potvrzení, či vyvrácení existence jasného kauzálního vztahu mezi ukazateli výkonnosti firem a jejich klasifikací z pohledu provozované ekonomické činnosti a velikosti dle prezentovaného počtu zaměstnanců. Pro potřeby ověření těchto dvou hypotéz je nutné pomocí vytvořené aplikace provedení základní finanční analýzy poměrových ukazatelů na základě veřejně přístupných účetních výkazů stovky českých firem zařazených v žebříčku Czech 100 Best [20]. Výstupem finanční analýzy bude matice dat sloužící jako podklad k potvrzení, či vyvrácení stanovených hypotéz prostřednictvím vícerozměrných statistických metod, a to pomocí shlukové analýzy.

Splnění hlavního cíle této práce lze dovodit z výstupů jednotlivých na sebe navazujících dílčích úkolů a cílů, kterými jsou:

- Ukotvení teoretických východisek a znalostí v oblasti finanční analýzy a vybraných statistických metod.
- Vyhledat zdroje potřebných informací a shromáždit všechny potřebné dokumenty a data.
- Navrhnout vhodnou metodiku zpracování dat, podle níž se bude postupovat při digitalizace účetních závěrek, tak aby bylo umožněno jejich strojové zpracování.
- Navrhnout strukturu úložiště, které bude uchovávat údaje extrahované z jednotlivých účetních výkazů. Extrakce informací bude prováděna prostřednictvím vytvořené aplikace, která bude celý proces maximálně automatizovat.
- Vytvořit aplikaci, jenž na základě informací získaných z účetních závěrek vypočte definované rozdílové finanční ukazatele.
- Podrobit objekty (firmy) na základě jejich atributů (vypočtených finančních ukazatelů) vhodně nastaveným metodám shlukové analýzy.
- Podrobit hypotézy testu, zda lze potvrdit, nebo vyvrátit, že existuje vazba mezi ukazateli výkonnosti firem a jejich klasifikací z pohledu provozované ekonomické činnosti a velikosti.

Práce je rozdělena do tří hlavních částí. První část shrnuje teoretická východiska, druhá část je analytické a třetí pak návrhová.

Teoretická část práce se zabývá finančními výkazy zejména ve vztahu k právním nárokům, které jsou kladeny na jejich strukturu a způsob zveřejňování. Dále pak představuje základní metody finanční analýzy. Teoretické část obsahuje také představení technologií, které budou dále využity při tvorbě potřebných analytických aplikací a v neposlední řadě představuje elementární metody a postupy shlukové analýzy.

Analytická část se věnuje představení jednotlivých analyzovaných subjektů v rozsahu, který je nezbytný k úspěšnému splnění vytyčených cílů. Dále je pak věnován prostor analýze zdrojů informací o firmách a také dokumentům, které tyto firmy v souladu se zákonem zveřejnily. V závěru analytické části je představena metodika zpracování dat a samotný princip provedení finanční analýzy. Výstupem analytické části je datová matice s vypočtenými finančními ukazateli.

Návrhová část práce je věnována čistě aplikaci metod a postupů shlukové analýzy na daných objektech a v samotném závěru je konstatováno, zda došlo k potvrzení nebo vyvracení formulovaných hypotéz.

1. TEORETICKÁ VÝCHODISKA PRÁCE

1.1 Účetní výkazy

K významným zdrojům vstupních dat pro zpracování finanční analýzy patří účetní výkazy podniku, tedy rozvaha, výkaz zisků a ztráty, přehled o peněžních tocích (cash flow), přehled o změnách vlastního kapitálu a příloha k účetní závěrce. Mnoho pro analytika cenných informací lze získat také ze samotné výroční zprávy. Informace lze dále čerpat od vrcholového vedení podniků, z interních zpráv, z firemních statistik, od auditorů, či z oficiálních ekonomických statistik [1].

S daty z účetních výkazů je třeba zacházet obezřetně, neboť jejich vypovídací schopnost je dána zejména účelem zpracování výkazů. Jsou zpracovávány pro účely účetní a daňové a nemusí tudíž vždy obsahovat data, které věrně zobrazují ekonomickou realitu podniku [1].

1.1.1 Právní minimum

Nejpodstatnější právní normy a zákony týkající se účetnictví a účetních výkazů:

1. **Zákon o obchodních společnostech a družstvech (zákon o obchodních korporacích) č. 90/2012 Sb. v platném znění** – stanovuje základní povinnosti všech podnikatelů vést účetnictví způsobem stanoveným zvláštním právním předpisem (zákon č. 563/1991 Sb., o účetnictví) [2] [1].
2. **Zákon č. 563/1991 Sb. v platném znění**, o účetnictví upravuje rozsah vedení účetnictví, účetní doklady, zápisy a knihy; stejně tak účetní závěrku; oceňování a inventarizaci majetku atp. [3] [1]

Paragrafy týkající se zpracovávání účetních výkazů:

§ 4, odst. 8: „Účetní jednotky jsou povinny dodržovat při vedení účetnictví zejména směrnou účtovou osnovu, uspořádání a označování položek účetní závěrky a konsolidované účetní závěrky, obsahové vymezení těchto, účetní metody, podmínky předávání a přebírání účetních záznamů a ostatní podmínky vedení účetnictví stanovené prováděcími právními předpisy. Prováděcí právní předpisy pro jednotlivé skupiny účetních jednotek upraví“

§ 7, odst. 1: „Účetní jednotky jsou povinny vést účetnictví tak, aby účetní závěrka byla sestavena na jeho základě srozumitelně a podávala věrný a poctivý obraz předmětu účetnictví a finanční situace účetní jednotky tak, aby na jejím základě mohla osoba, která tyto informace využívá (dále jen „uživatel“), činit ekonomická rozhodnutí.“

§ 8, odst. 1: „Účetní jednotky jsou povinny vést účetnictví správné, úplné, průkazné, srozumitelné, přehledné a způsobem zaručujícím trvalost účetních záznamů.“

§ 18, odst. 1: „Účetní jednotky sestavují v případech stanovených tímto zákonem účetní závěrku. Účetní závěrka je nedílný celek a tvoří ji

a) rozvaha (balance),

b) výkaz zisku a ztráty,

c) příloha, která vysvětluje a doplňuje informace obsažené v částech uvedených pod písmeny a) a b), zejména naplněním § 7 odst. 3 až 5 a §19 odst. 5 a 6.“ [3]

1.1.2 Povinnost zveřejnění účetní závěrky

Subjekty zapsané v obchodním rejstříku mají povinnost v souladu se zákonem č. 90/2012 Sb. § 436 každoročně zveřejnit účetní závěrku jejím předáním rejstříkovému soudu. Účetní závěrka se vždy zveřejňuje v rozsahu, v jakém byla sestavována. Ve zjednodušeném rozsahu účetní závěrku zveřejňují účetní jednotky, které nejsou povinny mít svou závěrku ověřenou auditorem. S výjimkou akciových společností, které sestavují účetní závěrku vždy v plném rozsahu [4].

Na základě ustanovení obchodního zákoníku (č. 563/1991 Sb. § 38k) je podnikatel povinen předložit rejstříkovému soudu veškeré listiny bez zbytečného odkladu od vzniklé skutečnosti. Účetní jednotky, které mají povinnost mít svou účetní závěrku ověřenou auditorem ji zveřejňují spolu s výroční zprávou až po ověření auditorem a schválení valnou hromadou ve lhůtě 30 dní. Ovšem nejpozději do konce následujícího účetního období, a to již bez ohledu na to, zda byly záznamy řádným způsobem schváleny [4].

Za nesplnění povinnosti založení účetní závěrky do sbírky listin obchodního rejstříku může být uložena pokuta do výše 3 % hodnoty aktiv celkem. Účetní závěrky se do sbírky listin zasílají v čistě elektronické podobě (datová schránka, e-mail, elektronická podatelna) [4] [5].

Technická specifikace digitalizace pro sbírku listin vyhotovená Odborem informatiky Ministerstva spravedlnosti ČR definuje 6 základních technických předpokladů pro bezproblémové zařazení poskytnuté listiny do Sbírk listin [5].

1. Každá listina musí být tvořena právě jedním PDF souborem (listiny tedy nelze dělit, ani slučovat).
2. Listiny nesmí obsahovat informace, které nemohou být de zákona zveřejněny
3. Datová velikost dokumentů je limitována 150 kB na 1 stránku listiny.
4. Je doporučeno, aby PDF dokument byl vytvořen převodem z textových dokumentů, a nikoliv z vložených grafických obrazů.
5. Listiny lze ověřit zaručeným elektronickým podpisem a nastavit jejich zabezpečení; v žádném případě však není možné omezit je heslem, či snížit plnohodnotné možnosti zobrazení a tisku.
6. Soubory do celkové velikosti 10 MB mohou být podány elektronicky; větší soubory je nutno doručit (poštou, či osobně) na datovém nosiči [5].

1.1.3 Rozvaha

Rozvaha nám podává informaci o tom, jaký majetek podnik vlastní a z jakých zdrojů je majetek financován. Rozvaha se vždy sestavuje ke konkrétnímu dni a musí být splněna základní bilanční rovnice (aktiva se rovnají pasivům) [1] [6].

Aktiva zobrazují majetkovou strukturu podniku. Základním hlediskem jejich členění je především doba upotřebitelnosti, případně rychlost konverze v peněžní prostředky. Základní struktura aktiv v rozvaze:

- A Pohledávky za upsaný základní kapitál
- B Dlouhodobý majetek
- C Oběžná aktiva
- D Časové rozlišení [1]

Pasiva zachycují finanční strukturu podniku, obsahuje tedy zdroje financování podnikového majetku [1].

Základní struktura pasiv v rozvaze se skládá z:

- A Vlastní kapitál
- B Cizí zdroje
- C Časové rozlišení [1]

1.1.4 Výkaz zisku a ztráty

Výnosy, náklady a výsledek hospodaření tvoří obsah výkazu zisku a ztráty.

Výnosy jsou peněžní částky, jenž podnik získal z veškerých činností za účetní období. Naproti tomu náklady představují peněžní částky, které podnik účelně vynaložil na získání výnosů v daném účetním období. Výsledek hospodaření je pak tvořen rozdílem mezi celkovými výnosy a celkovými náklady. Pokud vyjde výsledek hospodaření záporný podnik hospodařil se ztrátou, v opačném případě vykázal zisk [1] [6].

1.1.5 Přehled o peněžních tocích

Výkaz zisku a ztráty zaznamenává rozličné kategorie výnosů, nákladů a zisku v okamžiku jejich vzniku, nebere však ohled na to, zda skutečně vznikne reálný peněžní příjem nebo výdaj. Tento obsahový a časový nesoulad odstraňuje sledování cash flow, které bezprostředně souvisí s finančním řízením a zajišťováním likvidity podniku. Výkaz cash flow objasňuje přírůstky a úbytky peněžních prostředků a také důvody, které k těmto pohybům vedly.

Výkaz cash flow se zpravidla člení na:

1. oblast provozní činnosti
2. investiční oblast
3. oblast externího financování [1]

1.1.6 CZ-NACE

NACE je standardní klasifikací ekonomických činností v rámci Evropské unie užívanou od roku 1970. Český význam zkratky znamená statistickou klasifikaci ekonomických činností. Dělení ekonomických činností funguje tak, že každé jednotce vykonávající

ekonomickou činnost lze přiřadit konkrétní NACE kód, čímž se vytváří rámec pro získávání statistických dat o činnostech v mnoha ekonomických oblastech. CZ-NACE je pak národní verze číselníku NACE [19].

Číselník sekcí (první úrovně) NACE:

- sekce **A** – zemědělství, lesnictví a rybářství
- sekce **B** – těžba a dobývání
- sekce **C** – zpracovatelský průmysl
- sekce **D** – výroba a rozvod elektřiny, plynu, tepla a klimatizovaného vzduchu
- sekce **E** – zásobování vodou; činnosti související s odpadními vodami, odpady a sanacemi
- sekce **F** – stavebnictví
- sekce **G** – velkoobchod a maloobchod; opravy a údržba motorových vozidel
- sekce **H** – doprava a skladování
- sekce **I** – ubytování, stravování a pohostinství
- sekce **J** – informační a komunikační činnosti
- sekce **K** – peněžnictví a pojišťovnictví
- sekce **L** – činnosti v oblasti nemovitostí
- sekce **M** – profesní, vědecké a technické činnosti
- sekce **N** – administrativní a podpůrné činnosti
- sekce **O** – veřejná správa a obrana; povinné sociální zabezpečení
- sekce **P** – vzdělávání
- sekce **Q** – zdravotní a sociální péče
- sekce **R** – kulturní, zábavní a rekreační činnosti
- sekce **S** – ostatní činnosti
- sekce **T** – činnosti domácnostní jako zaměstnavatelů; činnosti domácnostní produkujících blíže neurčené výrobky a služby pro vlastní potřebu
- sekce **U** – činnosti extraterritoriálních organizací a orgánů [21]

1.1.7 Klasifikace firem dle počtu zaměstnanců

Velikost podniku určujeme na základě nejrůznějších kvantitativních veličin, jedním z nejčastějších ukazatelů je počet zaměstnanců [22].

Číselník kategorizace firem podle počtu zaměstnanců (metodika OECD)

- kód **000**: neuvedeno
- kód **110**: bez zaměstnanců
- kód **120**: 1 - 5 zaměstnanců
- kód **130**: 6 - 9 zaměstnanců
- kód **210**: 10 - 19 zaměstnanců
- kód **220**: 20 - 24 zaměstnanců
- kód **230**: 25 - 49 zaměstnanců
- kód **240**: 50 - 99 zaměstnanců
- kód **310**: 100 - 199 zaměstnanců
- kód **320**: 200 - 249 zaměstnanců
- kód **330**: 250 - 499 zaměstnanců
- kód **340**: 500 - 999 zaměstnanců
- kód **410**: 1000 - 1499 zaměstnanců
- kód **420**: 1500 - 1999 zaměstnanců
- kód **430**: 2000 - 2499 zaměstnanců
- kód **440**: 2500 - 2999 zaměstnanců
- kód **450**: 3000 - 3999 zaměstnanců
- kód **460**: 4000 - 4999 zaměstnanců
- kód **470**: 5000 - 9999 zaměstnanců
- kód **510**: 10000 a více zaměstnanců [23]

1.2 Finanční analýza

Cílem finanční analýzy je komplexní zhodnocení finanční situace podniku, neboť průběžná znalost finanční situace firmy umožňuje manažerům činit správná rozhodnutí při řízení firmy, a to jak dlouhodobém, tak krátkodobém. Poměrně obtížná je v oboru situace externích analytiků, který není v přímém spojení s analyzovaným podnikem a musí se tak spokojit pouze s veřejně dostupnými údaji. Tato data jsou často agregovaná a neúplná [6] [1].

Při postupu analýzy z pohledů externího zpracování je podstatným krokem získání informací o samotné společnosti. Zdrojem těchto dat jsou zpravidla výroční zprávy publikované ve sbírce listin obchodního rejstříku [6] [1].

1.2.1 Ukazatele finanční analýzy

Absolutní ukazatele finanční analýzy jsou většinou používány k analýze vývojových trendů v časových řadách (horizontální analýza) a k procentnímu rozboru jednotlivých komponent (vertikální analýza) [1].

Rozdílové ukazatele finanční analýzy se užívají za účelem analýzy a řízení finanční situace podniku, kdy největší pozornost je kladena na jeho likviditu [1].

Poměrové ukazatele finanční analýzy

Poměrové ukazatele finanční analýzy tvoří základní nástroj finanční analýzy. Tato rodina ukazatelů nám umožňuje získat rychlou představu o finanční situaci podniku. Podstatou jejich tvorby je kladení do poměru různé položky z účetních výkazů analyzované společnosti [1].

V běžné praxi se užívá několik základních ukazatelů, ty lze dělit do skupin podle jednotlivých oblastí hodnocení hospodaření a finančního zdraví podniku:

- **Analýza zadluženosti, majetkové a finanční struktury**

Tato skupina ukazatelů slouží jako indikátor výše rizika, jež podnik nese za dané struktury pasív.

Mezi tyto ukazatele patří například celková zadluženost, míra zadluženosti, doba splácení dluhů, úrokové krytí a další [6] [1].

- **Analýza likvidity**

Vyjadřuje schopnost podniku hradit závazky. Vzorce jsou sestaveny tak, že číselník je vždy tvořen prostředky, jenž je možno užít k platbě a ve jmenovateli jsou vyčísleny závazky, které je nutno uhradit.

Rozlišujeme tři stupně likvidity: běžnou, pohotovou a okamžitou [6] [1].

- **Analýza rentability**

Jde o měřítko schopnosti dosahovat zisku za použití investovaného kapitálu. Jde o formu vyjádření míry zisku (schopnosti podniku tvořit nové zdroje).

Nejčastěji používanými ukazateli jsou rentabilita tržeb, rentabilita celkového kapitálu, rentabilita vlastního kapitálu, rentabilita investovaného kapitálu a rentabilita úplného kapitálu [6] [1].

- **Řízení aktiv (ukazatele aktivity)**

Ukazatele aktivity nám umožňují zjistit, zda podnik využívá efektivně vložené prostředky neboli zda je přiměřená velikost jednotlivých druhů aktiv v poměru k hospodářským aktivitám podniku.

Do této kategorie ukazatelů spadá obrat aktiv, obrat dlouhodobého majetku, doba obratu zásob, doba obratu pohledávek, doba obratu závazků a další [6] [1].

- **Ukazatele kapitálového trhu**

Užívají se v případech, kdy je analyzovaný podnik obchodován na kapitálovém trhu. Mezi tyto ukazatele patří zisk na akcii, Cash flow na akcii, P/E Ration atp [1].

1.2.2 V podnikatelské praxi nejčastěji užívané finanční ukazatele

Dle šetření, v němž bylo zastoupeno široké spektrum podniků, z let 2009 až 2010 využívá až 87 % podniků nástroje finanční analýzy [1].

Zdaleka nejčastěji užívaným finančním ukazatelem je velikost tržeb, potažmo jejich rentabilita. Dále pak jde o ukazatele čistého zisku, zisku před zdaněním, ukazatele využívající Cash flow a rentabilitu nákladů [1].

Průměrně užívanými ukazateli (v rozpětí 46 % - 35 % zastoupení mezi sledovanými podniky) jsou rentabilita vlastního kapitálu, rentabilita aktiv a rentabilita investovaného kapitálu [1].

1.3 Digitalizace dat a ETL

1.3.1 Optické rozpoznávání znaků (OCR)

Jedná se o metodu, které se užívá v případech, kdy je nutné, aby informace byla čitelná jak pro člověka, tak pro stroj. OCR za pomoci scanneru umožňuje digitalizaci tištěných

textů, nebo převod textů uchovaných v digitálních obrázcích do normálního počítačového textu. Pokud je zdroj pro OCR zpracování již v digitálním formátu odpadá část zahrnující digitalizaci dokumentu (získání obrazu). Výsledek výstupu je samozřejmě velmi závislý na kvalitě vstupních dokumentů. Čím vyšší budou platit omezení pro vstupní data, tím vyšší přesnosti bude systém dosahovat [7].

Historie prvních OCR systémů se datuje již do roku 1954, kdy tato zařízení uměla převádět ručně psané obchodní reporty na děrné štítky a s těmi tak bylo možné dále pracovat v počítači. V současné době užívané systémy se řadí ke čtvrté generaci OCR. Ta je charakteristická svou dostupností, a tedy masovou rozšířeností. Základním softwarovým OCR balíčkem je dnes vybaveno nepřeberné množství zařízení včetně v současné době prodávaných scannerů [7].

Princip OCR

Typický OCR systém sestává z několika komponent:

- **Získání obrazu**

Obraz je získáván pomocí jakéhokoliv zařízení schopného digitalizace dat (skener, fotoaparát, kamera). Tento krok odpadá, pokud máme zdroje již digitalizované, ale nejsou ve strojově čitelné podobě [7].

- **Předzpracování obrazu**

V tomto kroku se snažíme kompenzovat zhoršenou kvalitu digitální dat, která mohou obsahovat šum (odstraňuje se mediánovým filtrem), nebo mohou být pořízeny za špatných světelných podmínek (úprava jasu a kontrastu) [7].

- **Segmentace**

V tomto bodě se snažíme identifikovat objekty v obraze. Cílem je extrakce důležitých objektů (tedy těch nacházejících se v popředí) a jejich oddělení od pozadí obrazu. Nejjednodušší metodou je prahování, které vytvoří binární obraz. V tomto obraze je černým pixelům přiřazena hodnota 1. Vycházíme tak z předpokladu, že hledané oblasti mají podobnou úroveň jasu [7].

- **Popis obrazu**

Objekty, které segmentací získáme je nutné nějakým způsobem popsat a klasifikovat. Pro jejich popis OCR systémy užívají příznaky (jejich volba a kvalita zásadním způsobem ovlivňuje celkový výsledek výstupu). Obecně jde o přiřazení oblastí odpovídajícímu vzoru [7].

- **Klasifikace**

V rámci klasifikace již dochází k identifikaci samotných znaků, které jsme dostali po segmentaci a popisu obrazu.

Metoda nalezení shody

Nejjednodušší a nejstarší metodou je metoda nalezení shody, kdy se znak, který je třeba klasifikovat postupně srovnává s různými šablonami a hledá se největší shoda. Jde však o poměrně pomalý proces (kvůli počtu šablon) a flexibilita je také poměrně nízká.

Statistická klasifikace

V případě této metody se využívá pravděpodobnosti. Pro neznámý znak je spočítána pravděpodobnost, že znak patří do některé ze tříd, zařazen je následně do třídy s nejvyšší pravděpodobností výskytu.

Neuronové sítě

Velice často se v moderních systémech užívá umělá neuronová síť, kterou tvoří distribuovaný výpočetní systém [7].

ABBYY FineReader

Jde o komerční software. Aplikace ABBYY převádí texty z obrazových dokumentů do digitálních a editovatelných formátů (např. Microsoft Office dokumenty, CSV, HTML, PDF atp.), u nichž existují možnosti úprav a umožňuje tak uživatelům obsah těchto dokumentů například archivovat. FineReader podporuje široké spektrum jazyků. V současné verzi pokrývá 190 jazyků, včetně jakýkoliv jazykových kombinací a pro

48 jazyků disponuje vestavěnou kontrolou pravopisu. Aplikace je založena na technologii ABBYY OCR [8].

1.3.2 ETL proces

Pojmenování ETL proces vychází z jmen tří úkonů, které jsou v rámci tohoto procesu činěny:

- Extrakce (extraction) – výběr dat různými metodami
- Transformace (transformation) – ověření, čištění, časové označení a integrování dat
- Nahrávání (loading) – přemístění dat do datového skladu

V ETL procesu je hlavním cílem centralizace údajů a jejich náhraní do datového skladu. Na úrovni ETL se pracuje s daty, ze kterých se později stávají informace, tudíž práce s daty by měla být kvalitní a přesná [9].

Extrakce

Krok extrakce zahrnuje shromáždění dat z rozličných operačních a archivních zdrojů, přičemž data a soubory bývají obvykle heterogenního charakteru. Extrahovaná data chceme zahrnout do datového skladu. Systém extrakce by měl být navržen tak, aby nemohlo dojít k ohrožení zdrojového systému z pohledu výkonu (reakčních časů) [9].

Existuje několik způsobů, jak extrakci dat provést:

- Aktualizace oznámením

Aktualizace oznámením se používá v případech, kdy je zdroj schopen zaslat upozornění, že byl záznam změněn a popsat jeho změnu, v tomto případě se jedná o nejsnazší cestu, jak získávat data [10].

- Inkrementální extrakce

Některé systém nemusí být schopny poskytnout notifikaci, že k aktualizaci údajů došlo, ale stále jsou schopny identifikovat, které záznamy byly změněny a takové záznamy při extrakci označit. V dalších krocích pak už je třeba jen identifikovat změnu a zpropagovat ji do datového skladu [10].

- Úplná extrakce

Některé zdroje nejsou schopny identifikovat, která data byla změněna, nebo zda byla vůbec změněna, takže jediným způsobem, jak ze systému data vydolovat je úplná extrakce. V těchto případech se nabízí uchovávat kopii poslední extrakce a v dalších krocích pak najít data, která byla od poslední verze změněna a rozdíly opět zpropagovat do datového skladu [10].

Transformace

Obecně platí, že je lepší nemít žádná data než mít nekvalitní data. Krok transformace sestává z aplikace předem stanovených pravidel na zdrojová data, za účelem jejich transformace do podoby, která bude kompatibilní s cílovou databází. Z toho důvodu dochází k čištění dat a jejich transformaci. Samotná část transformace obvykle sestává z odstranění nejednoznačnosti údajů (různé značení pro stejné hodnoty), ošetření chybějících hodnot a duplicitních záznamů (v některých případech může být chybějící údaj velmi problematický) a ošetření konvence názvů a pojmů objektů (různé peněžní měny, formátování čísel a textových řetězců) [9].

Nahrávání

Poslední část ETL procesu sestává z nahrání dat do datového skladu. Zavádění dat by zpravidla mělo být plánované a maximálně automatizované, následně po nahrání dochází obvykle k indexování dat. Po dokončení nahrávání je nezbytné se ujistit, že proces nahrání dat do datového skladu proběhl korektně a výpočetní zdroje nebyly příliš přetěžovány. Obvyklou cílovou destinací nahrávání je databáze. Typicky se největší množství dat přenáší při prvním nahrávání, při dalších už pak dochází pouze k nahrávání rozdílových dat [9].

1.3.3 Technologické minimum

XML

Jedná se o značkovací jazyk (Extensible Markup Language) určený k organizaci údajů v textových dokumentech do takové podoby, aby jim byli schopni porozumět jak lidé, tak stroje [11].

Jednotlivé prvky (elementy) jsou organizovány do stromové struktury. Mezi počáteční a koncovou značkou je uveden obsah daného tagu, ten může být tvořen konkrétní hodnotou, ale také jedním nebo více prvky. Tag může nést další doplňující atributy [11].

HTML a CSS

HTML je hypertextový značkovací jazyk, který je určený zejména k tvorbě webových stránek. S jeho pomocí se vytváří dokumenty obsahující strukturovaný text, formuláře, hypertextové odkazy a různorodý multimediální obsah [11].

Struktura HTML dokumentu sestává z hlavičky a těla stránky. Syntaxe je podobná XML dokumentu a stejně tak se jedná o prostý textový dokument, který lze velmi jednoduše upravovat pomocí textového editoru [11].

CSS neboli kaskádové styly (Cascading Style Sheets) spatřily světlo světa v roce 1996 a nabouraly původní formu HTML, který počítala, že text bude vždy stylován vždy přímo na místě pomocí formátovacích tagů. S příchodem CSS tak došlo k oddělení obsahu (HTML) od formy (formátovacích prvků). CSS styly dnes nabízejí široké spektrum možností stylování, nicméně není zcela zaručeno totožné vykreslení všech užitých parametrů napříč nepoužívanějšími prohlížeči [11].

Javascript

Javascript je klientský programovací jazyk. Funguje na straně klienta (v prohlížeči) a běh skriptu je na klientské stanici závislý. Jde o jazyk užíván pouze v oblasti HTML stránek.

Jedná se o objektový jazyk. Vrcholem hierarchie objektů je window (okno). Dalšími objekty jsou Date, Math a string. Všechny ostatní objekty jsou pouze podobjekty window [12].

jQuery knihovna

jQuery je Javascriptový framework (jde o jeden soubor tzv. knihovnu). Po vložení tohoto souboru do HTML dokumentu se nám zpřístupní funkce této knihovny, což v praxi vede ke zjednodušení a zrychlení psaní určitých částí kódu, které se často opakují. V tomto frameworku jsou přepracovány selektory do podoby příbuznější CSS selektorům. Pro manipulaci s objekty existuje řada metod a funkcí, které mohou být spouštěny také při

definovaných událostech (najetí kurzorem, kliknutí a další). Framework má také zabudované kompletní AJAX řešení (asynchronní komunikaci se serverem) [13].

PHP

PHP (jde o zkratku pro PHP Hypertext Preprocessor) je skriptovací jazyk, jenž se vykonává na straně serveru. Pracuje uvnitř HTML dokumentu, čímž mu propůjčuje schopnost generování požadovaného obsahu. Nicméně aby se kód úspěšně provedl, je třeba jej uvodit speciálními tagy a soubor uložit se správnou koncovkou.

PHP dosahuje nejvyšší výkonnosti, pokud je provozován jako interní modul Apache serveru a je dostupné jako open-source. Ve skriptovacím jazyce PHP lze napsat rozsáhlé webové portály a tohoto jazyka využívá i četné množství open-source redakčních systémů. [12] [14].

MySQL

Jde o systém správy relačních databází. Komunikace mezi databází a serverem probíhá standardně v jazyce SQL (Structured Query Language). Jde o open-source řešení a velmi populární databázový systém, který je dostupný prakticky na všech webhostinzích v kombinaci s PHP a Apache. Tyto tři technologie spolu tvoří tzv. „Svatou PHP trojici“. MySQL systém se skládá ze samotného serveru, klienta (zpravidla phpMyAdmin) a dalších podpůrných nástrojů. [12] [15].

Lokální server (XAMPP)

Pokud chceme aplikace vyvíjet na lokálním počítači, tak je třeba pro platformu Windows nainstalovat všechny výše uvedené technologie (tedy Apache, PHP a MySQL). V současné době je však již k dispozici celá řada předpřipravených serverů, které lze v několika málo krocích nainstalovat a spustit [16].

XAMPP dokáže na lokálním počítači vytvořit off-line server se všemi potřebnými technologiemi (obsahuje PHP, Perl, MySQL server, phpMyAdmin pro správu databáze, SMTP server atp.) [16].

1.4 Shluková analýza

Pojem shluková analýza zastřešuje širokou skupinu metod, z nichž některé byly navrženy teprve v nedávné době a neustále jsou navrhovány postupy nové, případně dochází k modifikaci těch stávajících.

Základním cílem shlukové analýzy je zařadit objekty do shluků (skupin), přičemž dva objekty stejného shluku, by si měly být podobnější než dva objekty z dvou různých shluků. Objektem shlukové analýzy může být prakticky cokoliv (od rostlin po webové stránky). Aby mohla být podobnost stanovena a změřena, je nutné, aby byl každý objekt charakterizován pomocí vlastností [17].

1.4.1 Principy shlukové analýzy

Shlukování a klasifikace

Shlukování je jeden ze základních typů získávání znalostí bez ohledu na to, zda pro dosažení cíle použijeme statistických metod, či metod strojového učení. Rozlišujeme dva přístupy klasifikace objektů:

- Učení s učitelem

V tomto případě obsahuje vstupní datový soubor také informace o příslušnosti objektů do předem známých skupin. Cílem tohoto přístupu je vytvořit model, na jehož základě by následně mohly být objekty bez známé příslušnosti agregovány do známých skupin.

- Učení bez učitele

Při užití tohoto přístupu není předem známá příslušnost objektů a obvykle neznáme ani počet skupin, do nichž budeme objekty rozřazovat. Cílem této metody je pak klasifikovat všechny, do analýzy zahrnuté, objekty [17].

Vstupní datová matice

Předpokládáme určitý popis m -rozměrného pozorování objektu (proměnné). Počet těchto objektů je označován n . Matematické odvození některých statistických metod vychází ze vstupní matice. V této matici řádky představují vektory údajů o jednotlivých objektech

(definují objekt) a sloupce pak odpovídají daným proměnným. Vstupní matice má tedy rozměry $n \times m$ [17].

Shluky

Pojem se spojuje se shlukováním objektů na základě podobností vektorů, který je popisuje. Ovšem v oblasti vyhledávání informací lze hledat také shluky proměnných, kdy se preferuje hledání shluků v rámci atributů, a nikoliv celých objektů.

Další variantou je tvorba shluků na základě současného shlukování objektů i proměnných.

Shluky lze rozlišit na disjunktivní a překrývající se. Disjunktivní shlukování je spojeno s jednoznačným přiřazením objektu do skupiny. V případě překrývajícího se shlukování se neuzívá čistě dvouprvková množina hodnot (buďto do skupiny patří, nebo ne), ale každému objektu je přiřazena míra příslušnosti daného objektu ke každému ze shluků [17].

Míry podobnosti

Cílem disjunktivního shlukování je vytvořit takové skupiny objektů, aby objekt v jedné skupině byl co nejpodobnější objektům ve stejném shluku a současně co nejméně podobný objektům v jiných shlucích. Základní etapou shlukování je tudíž stanovení kritérií, na jejichž základě bude podobnost objektů zjišťována [18] [17].

V ideálním případě nabývají míry podobnosti hodnot od nuly (maximální rozdílnost) po jedničku (totožnost).

V případě kvantitativních dat užíváme pro vyjádření vztahu dvou objektů míry vzdálenosti (míra nepodobnosti). Tyto míry jsou založeny na prezentaci objektů v prostoru. Souřadnice tohoto prostoru představují jednotlivé proměnné [17].

Kritéria shlukování

Jakmile si stanovíme cíl shlukování, je důležité zvážit, jakých shluků chceme dosáhnout a na základě toho zvolit vhodný shlukovací algoritmus, jeho volba je ovlivněna také tím, zda chceme dosáhnout vytvoření disjunktivních, či překrývajících se shluků (fuzzy shlukování) [17].

Rozlišujeme dvě základní metody shlukování:

- ploché (někdy též označované jako nehierarchické, či metody rozkladu)

Tyto metody vytváří strukturu rozkladu a prvky vstupní matice se rozkládají do stanoveného počtu shluků podle zvolených kritérií. Zásadním problémem těchto metod bývá stanovení správného počtu shluků, tak aby výsledky měly reálnou interpretaci.

- hierarchické

Výsledkem těchto metod je vytvoření hierarchie skupin objektů (posloupnost). Metody se dále dělí na aglomerativní (dochází k postupnému shlukování, proces probíhá, dokud nevznikne jeden klastr) a divizivní (jeden klastr složený ze všech prvků se postupně rozděluje na podmnožiny objektů, dokud se v každém shluku nebude nacházet pouze jeden objekt) [18] [17].

Kritéria pro stanovení počtu shluků

Jedním z předpokladů metod nehierarchického shlukování je zadání počtu shluků. Obvykle ovšem při zkoumání struktury vstupních dat o počtu shluků nemáme konkrétní představu. K zjištění optimálního počtu shluků existuje celá řada přístupů [17].

Jednou z možností je provádět shlukování pro různé počty klastrů a ze získaných výsledků zjistit jejich optimální počet, neboť při prvotním zkoumání struktury v datech obvykle nemáme představu o konkrétním počtu shluků, avšak vždy je potřeba dbát na to, aby každý z uvažovaného počtu shluků měl reálnou interpretaci [17].

V některých případech můžeme využít také hierarchického shlukování a vykreslení dendrogramu, na jehož základě budeme schopni určit počet shluků pro ploché shlukování. V případě většího počtu objektů je však dendrogram pro určení počtu shluků prakticky nepoužitelný [17].

Pro stanovení počtu shluků lze využít také některý z testů statistických hypotéz. V případě metod založených na hustotě provádíme posloupnost testů, kde nulová hypotéza označuje, že počet shluků je menší nebo roven k , a alternativní hypotéza naopak říká, že počet shluků je větší. Pro tyto potřeby můžeme využít metod Hubertovy statistiky [17].

Předzpracování datového souboru

Před samotným analyzováním vstupní datové matice bychom měli věnovat pozornost předzpracování dat. Musíme posoudit, jestli je vhodné do analýzy zahrnout všechny proměnné. Také bychom měli identifikovat, jednotlivé typy proměnných, standardizovat je a zjistit jejich významnost pro analýzu. Podstatné je také identifikovat odlehlé objekty (velmi odlišné od ostatních), protože může dojít ke zkreslení výsledků. Při užití základních algoritmů vytváří takové objekty samostatné shluky a neodlehlé objekty vytváří jeden společný shluk. Pokrýt bychom měli také situace, kdy některý z atributů ve vstupní matici zcela chybí, protože některé z programů takové objekty z posuzování vyřazují [17].

Výběr proměnných

K výběru proměnných lze přistupovat z věcného hlediska, kdy výběr proměnných závisí na znalosti problematiky. Musíme určit proměnné důležité z hlediska posouzení podobnosti objektů.

Ze statistického hlediska pak některé metody vyžadují, aby ve vstupní matici zůstaly pouze statisticky nezávislé proměnné. V praxi může tato podmínka způsobovat problémy, protože při důsledné aplikaci by mohl zůstat jen velmi malý počet proměnných mezi nimiž není silná závislost [17].

Rozlišujeme tyto základní typy proměnných:

- nominální – u hodnot lze rozlišit, zda jsou různé, avšak nelze stanovit jejich pořadí (např. typ absolvované střední školy)
- ordinální – můžeme stanovit pořadí, ale nemůžeme určit, o kolik je jedna hodnota vyšší, či nižší než jiná (např. stupeň souhlasu s výrokem)
- kvantitativní - číselné hodnoty, u nichž můžeme určit o kolik, či kolikrát je jedna hodnota rozdílná od jiné [17]

1.4.2 Metody rozkladu

U tohoto typu shlukování je vždy vytvářen konkrétní (předem stanovený) počet shluků.

Metoda k -průměrů

Užívá se v případech, kdy datový soubor obsahuje pouze kvantitativní proměnné a slouží k jednoznačnému přiřazení objektu do shluku. Jedná se o iterativní optimalizační metodu [17].

Metoda vychází z počátečního rozdělení objektů do zadaného počtu (k) shluků na základě určení počátečních centroidů, jenž tvoří středy shluků. Ty mohou být zvoleny různými způsoby (například první objekty ze vstupní matice). Poté se zkoumají vzdálenosti jednotlivých objektů od zvolených centroidů, tak že se pro každý objekt spočte euklidovská vzdálenost. Objekt je přiřazen k centroidu, který mu je nejbližší. Po dokončení první iterace je pro každý shluk spočten nový centroid, jehož vektor získáme spočtením průměrných hodnot jednotlivých objektů ve shluku. Dále pak znova testujeme všechny objekty, zda mají nejbliž ke shluku, do něhož jsou zařazeny. Pokud se nejbližší centroid po přepočtení nachází v jiném shluku, je do něj objekt zařazen. Tyto kroky se dějí tak dlouho, dokud dochází k přeskupování objektů mezi shluky [18] [17].

Metoda je vhodná pro soubory s velkým počtem vstupních objektů. Nevýhodou je, že nemusí být vždy nalezeno nejoptimálnější řešení (může jít pouze o lokálně optimální řešení), hodnotné řešení se zpravidla získá pouze pokud se zvolí několik rozdílných počátečních rozdělení [18] [17].

Metoda k -medoidů

Tato metoda je také známá pod zkratkou PAM (Partitioning Around Medoids). Algoritmus je stejně jako u předchozí metody určen pro proměnné kvantitativního typu a jejich rozdělení do k shluků. Metoda spočívá ve zjištění medoidu každého shluku. Medoidem je konkrétní objekt ze shluku a jako počáteční medoid je zvolen takový objekt, aby součet vzdáleností ostatních objektů ve shluku byl co nejmenší. V rámci algoritmu se pak iterativně postupuje tak, že pokud má objekt nejbliž k vlastnímu medoidu, tak je ve shluku ponechán. V opačném případě je přesunut do shluku, k němuž má nejbliž. V rámci iterace jsou pak medoidy vypočteny minimalizací funkce, jenž je tvořena součtem vzdáleností objektu od medoidu v daném shluku [17].

1.4.3 R

R je programovací jazyk a prostředí, které slouží pro statistickou analýzu dat a jejich vizualizaci. R můžeme vnímat jako implementaci jazyka S pod svobodnou (GNU) licenci. Prostedí R nabízí široké možnosti statistických (lineární a nelineární modelování, klasické statistické testy, analýzy časových řad, klasifikace a shlukování) a grafických technik. Jeho velkou výhodou je snadná rozšiřitelnost pomocí dodatečných balíčků [25].

2. ANALYTICKÁ ČÁST

2.1 Analyzované společnosti

Seznam firem byl získán z veřejně dostupného žebříčku českých 100 nejlepších firem za rok 2015 (Czech 100 Best) [20]. Žebříček obsahuje název společnosti, představitele společnosti spolu s jeho funkcí a IČ.

Legenda k tabulce

Firmy jsou seřazeny dle pořadí, jak se umístily ve výsledkové listině.

Uvedené názvy společností odpovídají názvům, které měly v roce 2015, u některých od té doby došlo k přejmenování (například společnost RWE ČESKÁ REPUBLIKA A.S. se přejmenovala na innogy Česká republika a.s.).

Zdrojem identifikačních čísel je výsledková listina, nicméně ve dvou případech muselo dojít ke korekci, protože IČ nekorespondovalo s názvem společnosti.

Informace o předmětu podnikání (CZ-NACE) byly získány vesměs z účetních výkazů a výročních zpráv, kde je zpravidla uveden hlavní předmět činnosti. Sekundárně pak z informací Českého statistického úřadu. V tabulce je vždy zastoupen převažující předmět činnosti. Uváděn je kód činnosti. Číselník je k dispozici v kapitole CZ-NACE.

Informace o počtech zaměstnanců byly získány z Registru ekonomických subjektů Českého statistického úřadu. Uváděn je kód, který vychází z metodiky OECD kategorizace firem dle počtu zaměstnanců. Počet zaměstnanců se vztahuje ke konkrétní firmě (IČ) a nejde tak o hodnotu celé skupiny firem (v případech kdy v žebříčku figuruje mateřská společnost). Kompletní číselník je k dispozici v kapitole Klasifikace firem dle počtu zaměstnanců.

ŠKODA AUTO A.S.

IČ: 00177041

CZ-NACE: C

Zaměstnanci: 510

AGROFERT, A.S.

IČ: 26185610

CZ-NACE: G

Zaměstnanci: 320

KOOPERATIVA POJIŠŤOVNA, A.S.

IČ: 47116617

CZ-NACE: K

Zaměstnanci: 460

BEST, A.S.

IČ: 25201859

CZ-NACE: C

Zaměstnanci: 330

**KARLOVARSKÉ MINERÁLNÍ
VODY, A. S.**

IČ: 14706725

CZ-NACE: C

Zaměstnanci: 330

ČEPS, A.S.

IČ: 25702556

CZ-NACE: D

Zaměstnanci: 340

STUDENT AGENCY K.S.

IČ: 25317075

CZ-NACE: H

Zaměstnanci: 420

LINET SPOL. S R.O.

IČ: 00507814

CZ-NACE: C

Zaměstnanci: 340

ČEZ, A. S.

IČ: 45274649

CZ-NACE: D

Zaměstnanci: 470

MOUNTFIELD A.S.

IČ: 25620991

CZ-NACE: G

Zaměstnanci: 420

KPMG ČESKÁ REPUBLIKA, S.R.O.

IČ: 00553115

CZ-NACE: M

Zaměstnanci: 340

FINCENTRUM A.S.

IČ: 24260444

CZ-NACE: K

Zaměstnanci: 310

ALLIANZ POJIŠŤOVNA, A.S.

IČ: 47115971

CZ-NACE: K

Zaměstnanci: 410

VÍTKOVICE HOLDING, A.S.

IČ: 25816039

CZ-NACE: C

Zaměstnanci: 210

SIKO KOUPELNY A.S.

IČ: 26065801

CZ-NACE: G

Zaměstnanci: 340

**KAREL HOLOUBEK - TRADE
GROUP A.S.**

IČ: 25060996

CZ-NACE: G

Zaměstnanci: 310

S GROUP HOLDING, A.S.

IČ: 27334121

CZ-NACE: M

Zaměstnanci: 230

**AVE CZ ODPADOVÉ
HOSPODÁŘSTVÍ S.R.O.**

IČ: 49356089

CZ-NACE: E

Zaměstnanci: 410

TOP HOTELS GROUP A.S.

IČ: 26443392

CZ-NACE: G

Zaměstnanci: 310

LINDE GAS A.S.

IČ: 0011754

CZ-NACE: C

Zaměstnanci: 340

AAA AUTO INTERNATIONAL A.S.

IČ: 01759299

CZ-NACE: G

Zaměstnanci: 420

RWE ČESKÁ REPUBLIKA A.S.

IČ: 24275051

CZ-NACE: M

Zaměstnanci: 330

BOHEMIA SEKT, S.R.O.

IČ: 45358711

CZ-NACE: C

Zaměstnanci: 330

BONECO A.S.

IČ: 27403238

CZ-NACE: C

Zaměstnanci: 310

**ŘÍZENÍ LETOVÉHO PROVOZU
ČESKÉ REPUBLIKY, STÁTNÍ
PODNIK**

IČ: 49710371

CZ-NACE: H

Zaměstnanci: 410

EXCON, A.S.

IČ: 00506729

CZ-NACE: M

Zaměstnanci: 240

SPORTISIMO S.R.O.

IČ: 26194627

CZ-NACE: G

Zaměstnanci: 430

CZ LOKO, A.S.

IČ: 61672131

CZ-NACE: C

Zaměstnanci: 340

LENOVO TECHNOLOGY B.V.**ORGANIZAČNÍ SLOŽKA**

IČ: 27243869

CZ-NACE: M

Zaměstnanci: 230

KAPSCH TELEMATIC SERVICES**SPOL. S R.O.**

IČ: 27371531

CZ-NACE: J

Zaměstnanci: 310

LESY ČESKÉ REPUBLIKY, S.P.

IČ: 42196451

CZ-NACE: A

Zaměstnanci: 450

RENOMIA, A.S.

IČ: 48391301

CZ-NACE: K

Zaměstnanci: 330

OKIN GROUP, A.S.

IČ: 27449734

CZ-NACE: M

Zaměstnanci: 430

SCANSERVICE A.S.

IČ: 25648101

CZ-NACE: S

Zaměstnanci: 240

AGROTRADE, A.S.

IČ: 26729270

CZ-NACE: G

Zaměstnanci: 210

ZENOVA SERVICES S.R.O.

IČ: 25051865

CZ-NACE: N

Zaměstnanci: 310

GECO, A. S.

IČ: 63080737

CZ-NACE: G

Zaměstnanci: 420

KOH-I-NOOR HOLDING A.S.

IČ: 45797463

CZ-NACE: C

Zaměstnanci: 330

JANEK SPOL. S R.O.

IČ: 13642898

CZ-NACE: C

Zaměstnanci: 240

ALTA INVEST, A.S.

IČ: 60735244

CZ-NACE: G

Zaměstnanci: 120

ZVVZ GROUP, A.S.

IČ: 28128494

CZ-NACE: L

Zaměstnanci: 130

SYNOT W, A.S.

IČ: 25548832

CZ-NACE: R

Zaměstnanci: 310

OHL ŽS, A.S.

IČ: 46342796

CZ-NACE: F

Zaměstnanci: 410

G - TEAM A.S.

IČ: 45358028

CZ-NACE: C

Zaměstnanci: 310

DELOITTE ADVISORY S.R.O.

IČ: 27582167

CZ-NACE: M

Zaměstnanci: 340

ČESKÁ TELEVIZE

IČ: 00027383

CZ-NACE: J

Zaměstnanci: 450

PRAŽSKÁ ENERGETIKA, A.S.

IČ: 60193913

CZ-NACE: D

Zaměstnanci: 340

**ENERGETICKÝ A PRŮMYSLVÝ
HOLDING, A.S.**

IČ: 28356250

CZ-NACE: L

Zaměstnanci: 120

ERNST & YOUNG, S.R.O.

IČ: 26705338

CZ-NACE: M

Zaměstnanci: 340

MICROSOFT S.R.O.

IČ: 47123737

CZ-NACE: G

Zaměstnanci: 330

MARK2 CORPORATION**CZECH A.S.**

IČ: 25719751

CZ-NACE: N

Zaměstnanci: 420

ASE, S.R.O.

IČ: 47536586

CZ-NACE: F

Zaměstnanci: 240

CARBOUNION BOHEMIA, SPOL.**S R.O.**

IČ: 25692917

CZ-NACE: G

Zaměstnanci: 230

XAVERGEN, A.S.

IČ: 27460363

CZ-NACE: C

Zaměstnanci: 330

TEDOM, S.R.O.

IČ: 28466021

CZ-NACE: D

Zaměstnanci: 340

INTERCORA, SPOL. S R.O.

IČ: 47714018

CZ-NACE: F

Zaměstnanci: 210

HUAWEI TECHNOLOGIES**(CZECH) S.R.O.**

IČ: 27367061

CZ-NACE: G

Zaměstnanci: 330

MP KRÁSNO, A.S.

IČ: 25572890

CZ-NACE: C

Zaměstnanci: 340

ICZ A.S.

IČ: 25145444

CZ-NACE: J

Zaměstnanci: 330

ELTODO EG, A.S.

IČ: 45274517

CZ-NACE: C

Zaměstnanci: 340

LÁZNĚ BĚLOHRAD A.S.

IČ: 46504834

CZ-NACE: Q

Zaměstnanci: 330

MASO UZENINY PÍSEK, A.S.

IČ: 26419793

CZ-NACE: C

Zaměstnanci: 330

RUDOLF JELÍNEK A.S.

IČ: 49971361

CZ-NACE: C

Zaměstnanci: 310

IVECO CZECH REPUBLIC, A.S.

IČ: 48171131

CZ-NACE: C

Zaměstnanci: 430

SERVIS CLIMAX A.S.

IČ: 25352628

CZ-NACE: C

Zaměstnanci: 330

KASPER GROUP S.R.O.

IČ: 03600033

CZ-NACE: C

Zaměstnanci: 0

ČESKÉ DRÁHY, A.S.

IČ: 70994226

CZ-NACE: H

Zaměstnanci: 510

CRYTUR, SPOL. S R.O.

IČ: 25296558

CZ-NACE: C

Zaměstnanci: 310

**PRVNÍ BRNĚNSKÁ STROJÍRNA
VELKÁ BÍTEŠ, A.S.**

IČ: 00176109

CZ-NACE: C

Zaměstnanci: 340

KARSIT GROUP SE

IČ: 47455608

CZ-NACE: C

Zaměstnanci: 340

**STOKLASA TEXTILNÍ
GALANTERIE S.R.O.**

IČ: 25877666

CZ-NACE: G

Zaměstnanci: 320

GRADA PUBLISHING, A.S.

IČ: 48110248

CZ-NACE: J

Zaměstnanci: 240

ZLKL, S.R.O.

IČ: 47973943

CZ-NACE: C

Zaměstnanci: 320

**ČESKÁ ZEMĚDĚLSKÁ
UNIVERZITA V PRAZE**

IČ: 60460709

CZ-NACE: P

Zaměstnanci: 430

CONTIPRO A.S.

IČ: 60917431

CZ-NACE: C

Zaměstnanci: 330

DVOŘÁK - SVAHOVÉ**SEKAČKY S.R.O.**

IČ: 26013797

CZ-NACE: C

Zaměstnanci: 240

ALTECH, SPOL. S R.O.

IČ: 46344861

CZ-NACE: C

Zaměstnanci: 320

CIMEX INVEST S.R.O.

IČ: 27159868

CZ-NACE: M

Zaměstnanci: 240

PODZIMEK A SYNOVÉ S.R.O.

IČ: 46978194

CZ-NACE: F

Zaměstnanci: 310

EKOL, SPOL. S R.O.

IČ: 41600983

CZ-NACE: C

Zaměstnanci: 310

IMS-DRAŠNAR S.R.O.

IČ: 25273396

CZ-NACE: C

Zaměstnanci: 310

YVENTECH, S.R.O.

IČ: 03553906

CZ-NACE: M

Zaměstnanci: 210

MILETA A.S.

IČ: 45534403

CZ-NACE: C

Zaměstnanci: 330

HOLÍK INTERNATIONAL S.R.O.

IČ: 25322214

CZ-NACE: C

Zaměstnanci: 310

CAROLLINUM S.R.O.

IČ: 64583350

CZ-NACE: G

Zaměstnanci: 230

HIT OFFICE GROUP A.S.

IČ: 55815441

CZ-NACE: G

Zaměstnanci: 0

**VYSOKÁ ŠKOLA TECHNICKÁ A
EKONOMICKÁ V ČESKÝCH
BUDĚJOVICÍCH**

IČ: 75081431

CZ-NACE: P

Zaměstnanci: 320

IDEA RS S.R.O.

IČ: 28356586

CZ-NACE: M

Zaměstnanci: 210

NEVŠÍMAL, A. S.

IČ: 25727133

CZ-NACE: F

Zaměstnanci: 240

INOS ZLIČÍN, A.S.

IČ: 25725459

CZ-NACE: G

Zaměstnanci: 240

K+K AUTOSERVIS CZECH S.R.O.

IČ: 27401197

CZ-NACE: G

Zaměstnanci: 240

KOBIT, SPOL. S R.O.

IČ: 44792247

CZ-NACE: C

Zaměstnanci: 310

INEKON GROUP, A.S.

IČ: 63998076

CZ-NACE: G

Zaměstnanci: 240

EXPLOSIA A.S.

IČ: 25291581

CZ-NACE: C

Zaměstnanci: 340

ČESKÝ PORCELÁN, A.S.

IČ: 00174238

CZ-NACE: C

Zaměstnanci: 330

C.O.T. MEDIA S.R.O.

IČ: 25098853

CZ-NACE: M

Zaměstnanci: 210

LAC, S.R.O.

IČ: 46903470

CZ-NACE: C

Zaměstnanci: 320

INSIGHT HOME, A.S.

IČ: 27942147

CZ-NACE: L

Zaměstnanci: 120

FOREZ S.R.O.

IČ: 64788342

CZ-NACE: C

Zaměstnanci: 340

PRIMOCO UAV SE

IČ: 03794393

CZ-NACE: C

Zaměstnanci: 130

2.2 Zdroje dat

Podkladová data (účetní závěrky a výroční zprávy), který jsou nezbytná pro zpracování základních metod finanční analýzy byly získávány z Veřejného rejstříku a sbírky listin Ministerstva spravedlnosti. Výjimku tvořily pouze dva analyzované subjekty z prostředí terciárního vzdělávání, které neuveřejňují účetní závěrky ve Veřejném rejstříku, avšak podle zákona č. 111/1998 Sb. o vysokých školách mají tyto subjekty povinnost veřejně zpřístupnit výroční zprávu o hospodaření vysoké školy. Zpravidla se tak děje prostřednictvím jejich webové prezentace.

Byla stahována data za kalendářní rok 2015, případně hospodářský rok, jehož významná část byla zasazena právě do kalendářního roku 2015. V sedmi případech pak byly užity výkazy čistě za rok 2014. Důvodem volby tohoto období byla motivace, aby naprostá většina společností již účetní závěrky měla zveřejněny a výkazy byly co nejaktuálnější.

2.2.1 Kvalita a dostupnost podkladů

Všechny níže popsané přehledy a statistiky popisují situaci k 28. 2. 2017 na vzorku sta vybraných společností. Všechna popisovaná data jsou prezentována prostřednictvím grafů v příloze na straně číslo I.

Ze vzorku 100 společností v roce 2017 nemají 4 společnosti zveřejněnou účetní závěrku za výše uvedená účetní období, přičemž tři společnosti nemají zveřejněnou jedinou účetní závěrku za celé období fungování společnosti a jedna z nich má poslední zveřejněnou závěrku z roku 2013. V případě těchto společností pravděpodobně nedochází k naplnění zákonné povinnosti zveřejňovat účetní závěrku jejím předáním rejstříkovému soudu. Tyto společnosti budou z dalších analýz logicky vynechány.

Celých 96 % společností mělo účetní závěrku za léta 2014 nebo 2015 zveřejněnou, ovšem velmi problematickým bodem je samotná kvalita poskytovaných dat. Kvalita dat má zásadní vliv na rychlost a možnost automatizace zpracování většího množství

elektronicky dostupných dokumentů. U nekvalitních dat musí docházet k následnému ověřování, zda převedené údaje skutečně odpovídají výchozím podkladům. Pokud je výchozí dokument horší kvality, velmi často data stoprocentně nekorespondují a musí docházet k manuálním úpravám. Čas nutný na zpracování takového dokumentu je až čtyřnásobný oproti dokumentům zpracovaných a publikovaných v souladu se zásadami publikovanými Ministerstvem spravedlnosti ČR (viz kapitola 1.1.2). Dokumenty připravené dle těchto zásad v případě dobře nastaveného procesu mohou obejít i bez lidského zásahu.

Dle technické specifikace Odboru informatiky Ministerstva spravedlnosti ČR pro digitalizaci sbírky listin je doporučeným formátem pro publikování zpráv PDF dokument vytvořený převodem z textových dokumentů, a nikoliv vložením grafických obrazců. Navzdory tomuto doporučení je 65 % všech dokumentů ve formě scanu. Týká se zejména účetních výkazů, které jsou pro datovou analýzu klíčové. Výjimkou však nejsou ani scany celých výročních zpráv, což jak již bylo zmíněno výše velmi znesnadňuje automatizované zpracování takovýchto dokumentů a celý proces prodražuje. V množině subjektů s nascanovanými účetními závěrkami mělo 13 % zdrojových dokumentů takovou kvalitu, že jejich zpracování bylo velice problematické a neobešlo se bez rozsáhlejších lidských zásahů. Typickými vadami jsou vysoká zrnitost, nízké rozlišení dokumentů, rotační deformace dokumentů, uříznuté části dokumentů atp. Ojedinelé se vyskytovaly i takové případy, kdy dokument nebylo možné přečíst ani lidským okem. Ve výjimečných případech docházelo k rozdělení jednotlivých výkazů do více listin ve Sbírce, nebo například ochraně PDF dokumentů před extrakcí (strojovým čtením). V jednom případě byly účetní výkazy vyplněny rukou a následně nascanovány.

Dvě společnosti pak v informačním systému Českého statistického úřadu nemají uveden údaj o počtu zaměstnanců.

Výkaz peněžních toků, který je nepovinnou součástí účetní závěrky mělo zveřejněno pouze 43 % společností, tudíž ukazatele založené na peněžních tocích, nebudou součástí analýzy.

2.2.2 Do analýzy nezahrnuté společnosti

Z výše uvedených důvodů nedostupnosti potřebných podkladů do analýzy nejsou zahrnuty společnosti JANEK SPOL. S R.O., YVENTECH, S.R.O., K+K AUTOSERVIS CZECH S.R.O. a PRIMOCO UAV SE.

Vyjmuty byly také obě společnosti podnikající v oblasti pojišťovnictví. Podoba jejich účetních výkazů se řídí vyhláškou č. 502/2002 Sb. Ministerstva financí, tak aby odrážela specifika podnikatelské činnosti. Majetek je v pojišťovnách členěn do dvou odlišných skupin z hlediska účelu, jde o finanční umístění a ostatní majetek. Ve struktuře účetních výkazů těchto společností jsou významné obsahové rozdíly [24].

2.3 Proces zpracování dat

Celý proces zpracování dat sestává z několika samostatných kroků a sleduje strukturu klasického ETL procesu. Celý proces je po krocích zjednodušeně zobrazen pomocí vývojového diagramu v příloze na straně číslo II.

Základní kroky sestávají ze získání podkladů, jejich převodu do strojově čitelného formátu, následné zpracování tohoto dokumentu a jeho import do datového úložiště. Nad takto uloženými daty může proběhnout finanční analýza, jejíž výsledky jsou posléze uloženy do databáze.

2.3.1 Digitalizace podkladů

Získané dokumenty ve formátu PDF je nutné převést do textové podoby, která by umožnila automatizované zpracování takovýchto dat. K digitalizaci využijeme ABBYY Finereader, který umožňuje scanovat PDF dokumenty a následně je ukládat ve formátu HTML.

U nekvalitní podkladových dat je třeba u vybraných stránek provést korekci zešikmení, či úpravu jasu a kontrastu podkladového dokumentu. Samotná digitalizace účetních výkazů probíhá na pokyn uživatele, kdy daná oblast dokumentu je označena jako nositel informací ve formátu tabulky. Po naskenování dokumentu lze provést vizuální kontrolu přečtených dat, kdy program zvýrazňuje problematické znaky, které by nemusely zcela odpovídat předloze.

2.3.2 Datový editor

Datovým editorem rozumím aplikaci vytvořenou pomocí webových technologií (HTML, CSS, Javascript, PHP a MySQL). Vstupem této aplikace je soubor ve formátu HTML, který obsahuje v tabulce strukturovaná data účetních výkazů. Výstupem je dávka databázových SQL příkazů, které reprezentují vstupní soubor.

Proces zpracování souboru

Zpracování souboru probíhá ve čtyřech krocích.

1. V prvním kroku se nahrává vstupní datový soubor.
2. V druhém kroku se výkaz přiřadí k firmě, v případech, kdy v databázi ještě nefiguruje vytvoří se nový záznam a zvolí se rok, který výkazy reprezentují. Dále se pak u jednotlivých tabulek rozhoduje, jaká data reprezentují (aktiva, pasiva, výkaz zisků a ztrát a přehled cash-flow). U daných tabulek se pak volí sloupec, ve kterém se nacházejí klíče (tedy popis hodnoty) a samotné hodnoty.
3. Ve třetím kroku dochází ke zpracování dat dle zadaných vstupních parametrů, kdy se z daných tabulek vyexportují dvojice klíčů a hodnot. Ty se dále párují s výchozími hodnotami, které jsou v databázi přednastaveny a byly vytvořeny na základě vzorů účetních výkazů v plném rozsahu. Všechny nahrávané vstupy, tak mohou být jednoznačně spárovány s vlastnostmi, které popisují, ač je jejich označení v dokumentu libovolné. Stejně tak je možné vytvářet nové podpoložky hlavním kategoriím, aby struktura nahrávaných dat poskytovala úplný a nezkreslený obraz původního dokumentu. Pokud klíč párované položky odpovídá některému z předpřipravených záznamů spárují se položky automaticky bez zásahu uživatele. V opačném případě k párování položek dochází manuálně metodou drag & drop.
4. V posledním kroku dochází k validaci vstupů a data se odesílají do databáze. Validují a upravují se formáty čísel, dále pak se ošetřují vstupy s prázdnými hodnotami a aplikace uživatele upozorní na nepřirazené položky. Z každého nahrávání také vzniká log, pro případné pozdější dohledání chybových míst v importu.

Samoučící mechanismus

Časově nejnáročnějším krokem nahrávání dat do databáze je párování hodnot, protože u rozsáhlých výkazů se může jednat o více než 200 položek, které je třeba nahrát. Z důvodu nestejněho pojmenovávání hodnot napříč různými účetními výkazy se ukládají také jednotlivé unikátní popisky hodnot a párují se ke klíčům, jenž v databázi již figurují. Díky tomuto opatření se rozšiřuje množina prvků, z níž lze při párování vycházet a zvyšuje se tak pravděpodobnost, že některý z klíčů ke spárování bude odpovídat klíči, který již byl datovým editorem dříve zpracováván, což vede k úbytku zásahů nutných ze strany uživatele, a tedy také k úspoře času stráveného nad nahráním jednoho výkazu.

Nespornou výhodou řešení je také možnost rekonstruovat výkaz do původní podoby včetně originálních popisků hodnot.

Struktura datového úložiště

Kompletní schéma datového úložiště včetně relačních vazeb je znázorněno na diagramu v příloze na straně číslo III. Tabulky datového úložiště jsou normalizované. Schéma tabulky databáze nelze určit zcela jednoznačně, protože neobsahuje striktně jednu tabulku faktů. Normalizované databázové struktury nejsou zcela typické pro datové sklady, avšak tento návrh datového úložiště se jeví být pro danou problematiku nejvhodnější.

Centrální tabulkou je *rozvahy_statements*. Tabulka obsahuje záznamy o jednotlivých nahraných datových výkazech. Obsahuje i původní nahrávaný HTML dokument, včetně metadat ve formátu JSON, která byla získána na základě vstupů uživatele. Tato metadata vznikla při označování charakteru a struktury dat, která byla obsažena ve vstupním dokumentu.

Tabulka s výkazy odkazuje do tabulky *rozvahy_companys*. Tato tabulka nese informace o firmách. Dále čerpá z tabulky *rozvahy_cznace_list*, mezi těmito tabulkami však existuje vazba N:M, proto je vazba mezi tabulkami zprostředkována tabulkou *rozvahy_companys_cznace*.

Důležitou je tabulka *rozvahy_rows*, v níž se nacházejí jednotlivé záznamy (řádky) výkazů. Uložené hodnoty jsou popisovány tabulkou *rozvahy_labels*, kde jsou namapovány jednotlivé klíče uložených hodnot. Tato tabulka je pomocí relační vazby

rozšířena tabulkou *rozvahy_statement_groups*. Ta obsahuje původní hodnoty popisků ze vzorových účetních výkazů v plném rozsahu.

Poslední tabulkou s vazbou na centrální tabulku je *rozvahy_indicators*. Jsou v ní obsaženy informace o vypočítaných ukazatelích finanční analýzy, které vznikly jako výstup třídy pro analýzu uložených finančních výkazů. Tato třída je popsána v kapitole číslo 2.3.3. Údaje v tabulce s vypočítanými ukazateli jsou popsány pomocí dat uložených v tabulce *rozvahy_indicators_names*.

2.3.3 Provedení analýzy nad daty

Nad zpracovanými a v databázi uloženými daty provedeme finanční analýzu. Analýza sestává z dávkového vypočtení vybraných poměrových ukazatelů. Poměrové ukazatele byly zvoleny proto, že jejich výstupy lze srovnávat bez ohledu na velikost účetní jednotky. Z výstupu těchto výpočtů budu sestavovat datovou matici pro shlukovou analýzu.

Zvolené poměrové ukazatele a jejich výpočet

Zvolené ukazatele jsou z několika klíčových oblastí ve vztahu k hodnocení finanční situace podniku. Předmětem analýzy je oblast rentability, likvidity, aktivity a zadluženosti.

Referenční hodnoty pro jednotlivé ukazatele jsou uvedeny v legendě k výsledkům finanční analýzy v kapitole číslo 2.4 Výstup analýzy dat.

	Poměrový ukazatel	Výpočet
rentabilita	ROA (<i>rentabilita aktiv</i>)	$= \frac{\text{EBIT}}{\text{aktiva celkem}}$
	ROE (<i>rentabilita vlastního kapitálu</i>)	$= \frac{\text{VH po zdanění}}{\text{vlastní kapitál}}$
	ROS (<i>rentabilita tržeb</i>)	$= \frac{\text{VH po zdanění}}{\text{tržby}}$
	ROI (<i>rentabilita investovaného kapitálu</i>)	$= \frac{\text{VH po zdanění}}{\text{Dlouhodobý kapitál}}$

likvidita	Běžná likvidita	$= \frac{\text{oběžná aktiva}}{\text{krátkodobé závazky}}$
	Pohotová likvidita	$= \frac{\text{finanční majetek} + \text{krátkodobé pohledávky}}{\text{krátkodobé závazky}}$
	Okamžitá likvidita	$= \frac{\text{finanční majetek}}{\text{krátkodobé závazky}}$
	Obrat celkových aktiv	$= \frac{\text{tržby}}{\text{aktiva celkem}}$
aktivita	Obrat dlouhodobého majetku	$= \frac{\text{tržby}}{\text{dlouhodobý majetek}}$
	Doba obratu zásob	$= \frac{\text{zásoby} \times 360}{\text{tržby}}$
	Doba obratu pohledávek	$= \frac{\text{pohledávky} \times 360}{\text{tržby}}$
	Doba obratu závazků	$= \frac{\text{závazky} \times 360}{\text{náklady vynaložené na prodané zboží} + \text{výkonová spotřeba}}$
zadluženost	Míra zadluženosti	$= \frac{\text{cizí zdroje}}{\text{vlastní kapitál}}$
	Úrokové krytí	$= \frac{\text{EBIT}}{\text{nákladové úroky}}$

Tabulka 1 Vzorce pro výpočet finanční analýzy (Zdroj vzorců: [1])

Dokumentace skriptu zpracovávajícího analýzu

Hromadná finanční analýza je prováděna pomocí skriptu vytvořeného v jazyce PHP.

Třída: Analysis

Proměnné:

- *\$statementId*: integer
- *\$statementData*: array

Metody:

Public

- *__construct(integer \$statement_id)*: konstruktork
- *doAnalysis(boolean \$db = true)*: funkce řídící provedení analýzy; vrací *true* při úspěchu, jinak *false*

Protected

- *existAnalysis()*: vrací *true* pokud již byl výkaz analyzován, jinak *false*
- *getStatementValues()*: vrací pole záznamů výkazu
- *countROA()*: vrací hodnotu ROA, pokud ji lze z údajů vypočítat, jinak *false*
- *countROE()*: vrací hodnotu ROE, pokud ji lze z údajů vypočítat, jinak *false*
- *countROS()*: vrací hodnotu ROS, pokud ji lze z údajů vypočítat, jinak *false*
- *countROI()*: vrací hodnotu ROI, pokud ji lze z údajů vypočítat, jinak *false*
- *countCurrentRatio()*: vrací hodnotu ukazatele běžná likvidita, pokud ji lze z údajů vypočítat, jinak *false*
- *countAcidTestRatio()*: vrací hodnotu ukazatele pohotová likvidita, pokud ji lze z údajů vypočítat, jinak *false*
- *countAbsoluteLuquidityRatio()*: vrací hodnotu ukazatele okamžitá likvidita, pokud ji lze z údajů vypočítat, jinak *false*
- *countAssetTurnoverRatio()*: vrací hodnotu ukazatele obrat celkových aktiv, pokud ji lze z údajů vypočítat, jinak *false*
- *countFixedAssetsTurnover()*: vrací hodnotu ukazatele obrat dlouhodobého majetku, pokud ji lze z údajů vypočítat, jinak *false*
- *countInventoryTurnover()*: vrací hodnotu ukazatele doba obratu zásob, pokud ji lze z údajů vypočítat, jinak *false*
- *countReceivablesCollection()*: vrací hodnotu ukazatele doba obratu pohledávek, pokud ji lze z údajů vypočítat, jinak *false*
- *countDaysPayableOutstanding()*: vrací hodnotu ukazatele doby obratu závazků, pokud ji lze z údajů vypočítat, jinak *false*
- *countDebtToEquity()*: vrací hodnotu ukazatele míry zadluženosti, pokud ji lze z údajů vypočítat, jinak *false*

- *countInterestEarnedRatio()*: vrací hodnotu ukazatele úrokového krytí, pokud ji lze z údajů vypočítat, jinak *false*

Private

- *returnEBIT()*: vrací hodnotu EBIT (mezi výpočet)
- *returnSales()*: vrací hodnotu tržeb (mezi výpočet)
- *addToDatabase()*: ukládá vypočtené ukazatele do databáze; vrací *true* při úspěchu, jinak *false*

2.4 Výstup analýzy dat

Celkovým výstupem analytické části této diplomové práce. Vyplynuly ze všech v této části uváděných kroků a postupů. Data jsou prezentována v přehledné tabulce v přílohách na stranách číslo IV, V a VI. Data budou následně (v neformátované podobě) fungovat jako vstupní datová matice při návrhu shlukové analýzy. Tabulky vznikla jako export vypočtených finančních ukazatelů z databáze do textového souboru ve formátu CSV.

Zvýrazněné buňky vypovídají o tom, že daná hodnota je v mezích doporučovaných referenčních hodnot pro daný ukazatel. Referenční hodnoty jsou uvedeny v legendě níže.

Legenda k datům dle sloupců

1. **Obchodní název společnosti**
2. **CZ-NACE** převažující podnikatelské činnosti (dle číselníku v kapitole číslo 1.1.6)
3. Klasifikace firem dle **počtu zaměstnanců** (dle číselníku v kapitole číslo 1.1.7)
4. **Rentabilita aktiv** (doporučená hodnota minimálně 8 % dle [1])
5. **Rentabilita vlastního kapitálu** (doporučená hodnota větší než 10 % dle [1])
6. **Rentabilita tržeb** (doporučená hodnota větší než 10 % dle [1])
7. **Rentabilita investovaného kapitálu** (referenční hodnota není stanovena)
8. **Běžná likvidita** (doporučená hodnota v uzavřeném intervalu 1,5 – 2,5 dle [1])
9. **Pohotová likvidita** (doporučená hodnota v uzavřeném intervalu 1 – 1,5 dle [1])
10. **Okamžitá likvidita** (doporučená hodnota v uzavřeném intervalu 0,5 – 1 dle [1])
11. **Obrat aktiv** (doporučená hodnota neuvedena, avšak platí, že vyšší hodnota je lepší [1])

12. **Obrat dlouhodobého majetku** (doporučená hodnota neuvedena, avšak platí, že vyšší hodnota je lepší [1])
13. **Doba obratu zásob** (doporučená hodnota je dána oborovým průměrem, firma by se měla snažit dosahovat nižších hodnot [1])
14. **Doba obratu pohledávek** (doporučená hodnota je dána oborovým průměrem, firma by se měla snažit dosahovat nižších hodnot [1])
15. **Doba obratu závazků** (doporučená hodnota je taková, která je vyšší, než doba obratu pohledávek [1])
16. **Míra zadluženosti** (doporučená hodnota v uzavřeném interval $0,2 - 0,5$ dle [1])
17. **Úrokové krytí** (doporučená hodnota je vyšší než 5 dle [1], nula znamená, že firma neviduje žádné nákladové úroky; z důvodu pozitivního vnímání i nulových hodnot byly hodnoty odpovídající doporučeným hodnotám označeny jedničkou a všechny ostatní pak nulou)

3. NÁVRHOVÁ ČÁST

V návrhové části práce, která přímo navazuje na analytickou část, se budu zabývat ověřením hypotéz vyřčených v cílech práce.

Všechny následující postupy jsou aplikovatelné v prostředí R za užití knihovny s názvem *cluster* [29].

3.1 Vstupní datová matice

Vstupní datová matice byla získána jako výstup z provedených finančních analýz na shromážděných datech. V přehledné podobě je k dispozici v kapitole číslo 2.4, oproti zde prezentované verzi počítáme pouze s číselnými hodnotami (nikoliv s procenty) s přesností až na sedm desetinných míst.

Zdrojový soubor je ve formátu CSV, jako oddělovač je užit středník, soubor je uložen v kódování UTF-8. Import těchto dat a přiřazení do proměnné *data* v jazyce R provedeme následujícím příkazem:

```
data = read.csv([cesta k souboru], TRUE, ";", encoding="UTF-8")
```

Prvním argumentem je cesta k souboru, druhým pak popisujeme, zda soubor obsahuje první řádek se záhlavím. Třetí parametr definuje oddělovač jednotlivých sloupců a posledním parametrem definujeme kódování, v němž je soubor uložen. Definice kódování je důležitá, aby byly správně vykresleny všechny znaky obsahující diakritiku.

Ošetření prázdných vstupů

Při přípravě dat vhodných k provedení shlukové analýzy je důležité neopomenout ošetření chybějících hodnot, kdy dochází u objektů, které nejsou popsány všemi parametry k jejich vyloučení z datové matice. V případě naší datové matice, kdy její tvorba byla plně pod naši kontrolou víme, že jsou všechny objekty plně popsány, nicméně z procesního hlediska je vhodné tento krok neopomíjet.

```
data = na.omit(data)
```

Výběr oblasti hodnot

Pro další operaci s hodnotami musíme z matice všech dat vybrat pouze číselné hodnoty a oříznout tak všechny popisující sloupce. Výřez z matice do proměnné provedeme níže uvedeným příkazem, jehož parametry se pro jednotlivé zde uváděné operace mohou lišit, neboť popisky uzpůsobují potřebě a výřez tak nebude vždy nutné aplikovat na všechny tři první sloupce jako v uvedeném příkladu.

```
data.matrix = data[,-c(1,2,3)]
```

Příprava dat

Ze zběžného pohledu na datovou matici je zřejmé, že jsou data v různých měřících, tudíž bychom měli provést standardizaci dat. V knihovně *cluster*, je k dispozici funkce *daisy*, která umožňuje standardizaci dat provést voláním jediné funkce a také umožňuje například jednotlivým parametrům objektů přiřadit určité váhy. Nicméně v rámci maximální kontroly nad procesem standardizace provede jednotlivé kroky postupně.

Standardizaci provedeme dvěma vektory. Prvním vektorem je medián, který vypočteme pro každý ze sloupců. Tento střed bude následně odečten od každé hodnoty v daném sloupci. Další hodnotou, kterou vypočteme je střední absolutní odchylka od mediánu. Touto hodnotou pak bude dělen rozdíl původní hodnoty a středu. Výše popsané operace provedeme příkazem *scale*. Jde o obecnou funkci jazyka R, která na základě výše popsaných vstupních parametrů provede centrování a škálování sloupců.

```
medians = apply(data.matrix, 2, median)
```

```
mads = apply(data.matrix, 2, mad)
```

```
data.matrix = scale(data.matrix, center=medians, scale=mads)
```

Číslo dvě v prvním a druhém příkazu říká, že medián a odchylku počítáme pro sloupce, a nikoliv pro řádky, výstupem těchto dvou příkazů jsou tedy matice se shodným počtem sloupců, jako má výřez vstupní datové matice.

3.2 Hierarchické shlukování

Jako první metodu zvolíme hierarchickou metodu shlukování. Její výhoda tkví v tom, že je nám schopna nastínit, jak bude vypadat řešení plochého shlukování při různém počtu

shluků. Výstupem hierarchických metod shlukování je dendrogram, jenž na ose x zobrazuje všechny analyzované subjekty a na ose y je vyobrazena vzdálenost na níž došlo ke spojení jednotlivých shluků. Výstup by nám také měl posloužit jako vodítko při stanovení počtu shluků pro tvorbu nehierarchických modelů.

V našem případě použijeme aglomerativní shlukování, kdy jsou postupně shlukovány nejpodobnější objekty. Shlukování probíhá tak dlouho, dokud nejsou všechny objekty propojeny do jednoho shluku spojujícího všechny objekty ze vstupní datové matice. Algoritmus probíhá v cyklech, kdy se v prvním kroku vypočte asociační matice a dojde ke spojení dvou nejpodobnějších objektů na základě zvolené metody aglomerativního shlukování, následně dojde k přepočítání asociační matice, kdy spojené objekty již vystupují jako jeden a opět se spojují dva nejpodobnější objekty.

Prvním krokem, který musíme učinit při hierarchickém shlukování je výpočet matice vzdáleností (všechny párové rozdíly mezi jednotlivými subjekty), díky provedené standardizaci dat nyní v datovém souboru nemáme žádné proměnné smíšeného typu.

Matici vzdálenosti můžeme v jazyce R získat dvěma způsoby, buďto pomocí funkce *dist*, nebo *daisy*, která je součástí knihovny *cluster*. V našem případě užitíme funkci *dist* s euklidovskou metodou výpočtu vzdáleností.

```
data.dist = dist(data.matrix , method="euclidean")
```

Vypočtenou matici vzdálenosti užitíme jako vstupní parametr funkce *hclust*, alternativně lze užít funkci *agnes*. Funkce nám vypočte množiny dat, které jsou výstupem aglomerativního hierarchického shlukování. Tyto množiny nám budou dále sloužit jako podklad pro grafickou prezentaci dat pomocí dendrogramů. Použijeme metodu *ward.D2*, jde o aktualizovanou implementaci původní metody *ward*. Tato metoda je vhodná pro kvantitativní proměnné.

```
data.clusts = hclust(data.dist, method="ward.D2")
```

Vykreslení vypočtených množin provedeme následujícím příkladem:

```
plot(data.clusts, labels=data$Firma, main='Hierarchické  
shlukování')
```

Eliminace extrémních hodnot

Z prvního výstupu shlukování je zřejmé, že v datovém souboru jsou minimálně tři odlehlé subjekty s výrazně extrémními hodnotami, které graf značně zkreslují a neumožňují se výrazněji zaměřit na nižší stupně, neboť ostatní (neodlehlé) objekty vytvoří jeden velký shluk. Z toho důvodu v prvním kroku tyto tři subjekty v datovém souboru eliminujeme. Jedná se jmenovitě o Energetický a průmyslový holding, a.s. a S group holding, a.s., hodnoty rentability tržeb, doby obratu závazků a pohledávek jsou v absolutních hodnotách nestandardně vysoké. Další eliminovanou společností je HIT OFFICE GROUP A.S. u níž jdou do extrémních hodnot všechny tři ukazatele likvidity (firma má výrazný krátkodobý finanční majetek a prakticky nulové závazky). Všechny výrazně vychylující se hodnoty u těchto subjektů byly manuálně ověřeny, aby bylo možné vyloučit chybu při hromadném zpracování.

Bez těchto subjektů výstup hierarchické analýzy nabyl standardnější podoby, nicméně při zvýraznění shluků v datovém souboru stále existují subjekty, které tvoří jedno-objektový shluk. Tímto problémem se budu více zabývat při rozkladu výsledků.

3.2.1 Na základě činnosti subjektu

Rozdíl v grafech pro analýzu na základě činnosti a velikosti zkoumaných objektů se omezí pouze na popisky, tak aby se grafy staly přehlednějšími a lépe se v nich četlo. Graf s uvedeným oborem činnosti získáme níže uvedeným příkazem a k vidění je v příloze na straně číslo VII.

```
plot(data.clusts, labels=paste(data$NACE, data$Firma, sep=" - "),  
main='Hierarchické shlukování - CZ-NACE')
```

Orámování jednotlivých shluků bylo provedeno níže uvedeným příkazem, jako počet shluků bylo zvoleno číslo 16, což je počet unikátních činností dle první úrovně CZ-NACE, které jsou zastoupeny ve vstupní datové matici.

```
rect.hclust(data.clusts, k=16, border="red")
```

Výrazně lépe se však bude číst graf bez uvedených názvů firem, které pro nás v současné chvíli nejsou natolik podstatné, protože bude vizuálně velmi rychle zřejmé, jakým

způsobem se vytvořili shluky mezi činnostmi napříč objekty. Tento případ je vizualizován v příloze na straně číslo VIII.

Posouzení výstupu a rozklad výsledků

Posouzení výstupu

Z výsledků nelze jednoznačně prokázat přímý kauzální vztah uvedený v první hypotéze, tedy existenci přímého vztahu mezi činností subjektu a hodnotami ekonomických ukazatelů, i přesto však můžeme v datech identifikovat určité fenomény.

V levé části dendrogramu se vytvořily samostatné a značně se odlišující shluky. Oblast je tvořena zejména subjekty ze sekcí G (velkoobchod a maloobchod; opravy a údržba motorových vozidel) a M (profesní, vědecké a technické činnosti). Tyto dvě sekce mají často obsahují mateřské firmy velkých skupin, které pro celý koncern obvykle zajišťují pouze servisní a podpůrné činnosti (správa budov, facility management, právní a poradenské činnosti) a je tudíž problematické tyto společnosti klasifikovat z pohledu oboru podnikání a velikosti. V ojedinělých případech je problém s klasifikací podpořen i tím, že poskytují pouze konsolidované účetní výkazy, což se však týká pouze jednotek společností ve sledovaném vzorku. Velmi blízko této skupině je také společnost ČEZ, a. s. ze sekce D (výroba a rozvod elektřiny, plynu, tepla a klimatizovaného vzduchu).

V pravé části dendrogramu se vytvořily také samostatně stojící objekty, nicméně jejich zařazení již není tak snadné provést. Překvapující je podobnost mezi jedinou veřejnou vysokou školou v žebříčku a společností Vítkovice holding, a.s., ke skupině se následně přidá Česká televize a větev dotvoří společnost CIMEX INVEST s.r.o., která má jako převažující činnost v účetní závěrce uvedeny právní a účetní činnosti spadající do sekce M, avšak na webu se prezentuje jako společnost podnikající v nemovitostech (sekce L), z čehož je patrné, že struktura této větve je velmi různorodá.

V grafu je patrných ještě několik samostatných shluků, které jsou tvořeny pouze jedním subjektem. Jedná se objekty, které nesou výrazně vybočující hodnoty.

Jako poměrně velmi zajímavé hodnotím, že jediné dva subjekty v žebříčku, jejichž právní forma je státní podnik a operují v naprosto odlišných oborech činností jsou si natolik

podobné, že první shluk vytvořily právě spolu. Jedná se o Řízení letového provozu České republiky, státní podnik (sekce H) a Lesy České republiky, s.p. (sekce A).

Není bez zajímavosti, že jeden ze středových shluků je až na tři výjimky tvořen prakticky výhradně společnostmi ze zpracovatelského průmyslu, což je v kontextu celého dendrogramu situace poměrně výjimečná, avšak ačkoliv jsou společnosti z jedné sekce dle CZ-NACE, tak svou podnikatelskou aktivitou jsou široce rozkročeny od nápojářského průmyslu, přes strojírenství a automobilový průmysl až po potravinářství. Ovšem svou velikostí zhruba poloviční shluk, stojící nalevo od výše popisovaného, je tvořen vyjma jedné společnosti pouze firmami působícími v různých oborech činností, které nespadají pod zpracovatelský průmysl, což je vzhledem ke štědrému zastoupení společností ze sekce C situace jedinečná. Velmi podobně je strukturován shluk okolo dvou výše zmíněných státních podniků.

Rozklad výsledků

Pro potřeby bližšího a komfortnějšího prozkoumání výstupů tohoto aglomerativního shlukování můžeme využít některou z dalších analytických funkcí jazyka R. Jde například o rychlé získání textového údaje, kolik objektů je přiřazeno v jednotlivých skupinách, při různém počtu shluků. Tyto hodnoty vypočteme příkazem:

```
clusters_size = cutree(data.clusts, 16)
```

informace pak vypíšeme v přehledné struktuře tabulky příkazem:

```
table(clusters_size)
```

a získáme následující výstup:

```
clusters_size
```

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
17 43  1  2 10  1  5  3  1  1  1  1  1  1  1  1
```

z něhož je na první pohled zřejmá situace popisována v předchozím oddíle a to, že máme velké zastoupení shluků, které mají pouze jeden objekt. Bude tudíž zajímavé se v dalším kroku podívat na výřez dat, který bude tvořen pouze objekty z prvního, druhého, pátého a sedmého shluku.

Pomocí tabulkového výpisu můžeme dostat také strukturu dat v jednotlivých shlucích, což je v kontextu interpretace výstupu jedna z klíčových informací. Díky tabulkové struktuře dat, je jasné patrná skladba firem v jednotlivých shlucích na základě provozované činnosti. Tabulku získáme níže uvedeným příkazem a je k dispozici jako první příloha na straně číslo X.

```
table(clusters_size, data$NACE)
```

Velmi užitečnou funkcí na jejímž základě lze získat charakteristické znaky skupiny je funkce *aggregate*. S její pomocí jsme schopni zjistit například mediány jednotlivých ukazatelů v rámci skupiny. Jednak tak můžeme zjistit v jaké finanční kondici dle konkrétních ukazatelů firmy ve skupině jsou, v čem tkví jejich vzájemná podobnost a můžeme na základě toho také odhadovat, do které skupiny bychom při znalosti ekonomických ukazatelů mohli zařadit libovolný $n + 1$ subjekt. V druhé příloze na straně číslo X je k vidění výřez mediánů hodnot dle jednotlivých skupin shluků, který je výstupem níže uvedené funkce.

```
aggregate(data[, -c(1,2,3)], list(clusters_size), median)
```

Můžeme zjišťovat také extrémy jednotlivých shluků, v takovém případě stačí v dané funkci poslední argument nahradit například hodnotou *min*, či *max*. Tento výstup dat je svou povahou vhodnější pro datové soubory s menším počtem vstupních proměnných. Výhodou pro vyvozování hlubších závěrů je, pokud jsou proměnné na první pohled analyticky uchopitelné hodnoty a nejedná se o výsledek různých matematických operací.

3.2.2 Na základě velikosti subjektu

Pro snazší posouzení výsledků prostřednictvím různých sumarizačních tabulek a agregačních funkcí bude praktické, když jednotlivé kategorie sjednotíme do několika málo základních skupin velikosti firem dle počtu zaměstnanců. Počet shluků odvodíme od první číslice klasifikační skupiny. V praxi to znamená, že všechny podniky zařazené ve skupinách 310–340 budou nově součástí skupiny 300. Obdobně se bude postupovat u všech tříd. Důvodem je menší roztržitost dat vzhledem k velikosti datového souboru.

Graf klasifikující objekty na základě počtu zaměstnanců dostaneme níže uvedeným příkazem a objekty rozdělíme do šesti shluků. Tento počet vychází z existence šesti základních skupin velikosti firem dle počtu zaměstnanců. Výstup je k dispozici v příloze

na straně číslo IX, v tomto grafu jsou pro úplnost zachovány původní hodnoty zařazení do skupin na základě počtu zaměstnanců.

```
plot(data.clusts, labels=data$Zamestnanci, main='Hierarchické  
shlukování - dle velikosti')  
  
rect.hclust(data.clusts, k=6, border="red")
```

Posouzení výstupu a rozklad výsledků

Posouzení výstupu

V tomto konkrétním případě, kdy došlo oproti kapitole číslo 3.2.1 ke snížení počtu shluků se ještě výrazněji vyjevil problém objektů s extrémními hodnotami, které zásadním způsobem znemožňuje interpretaci výsledného dendrogramu, neboť vzniklo pět shluků čítající pouze jeden subjekt a pak jeden velký shluk, který zahrnuje necelých 95 % objektů. Na základě takového výstupu bohužel nelze činit žádné konkrétní závěry. Pravděpodobně bude nutné analyzovat, již v předchozí kapitole zmiňovaný, výřez dat, jenž zahrnuje právě většinu subjektů v jedno shluku.

Rozklad výsledků

Díky vykreslenému dendrogramu máme již poměrně přesnou představu, jak vypadají jednotlivá seskupení objektů při šesti základních shlucích, ovšem mohlo by být zajímavé, pozorovat, jak se bude měnit skladba subjektů v jednotlivých shlucích, když budeme počet shluků navyšovat až k šestnácti užitým v předcházející kapitole. Toho můžeme docílit níže uvedeným příkazem, který kombinuje již výše zmíněný tabulkový výpis obsahující informace o počtu objektů v jednotlivých shlucích s jednoduchým cyklem, který bude provádět zvyšování počtu shluků až do maximální stanovené úrovně.

```
cl_sizes = sapply(6:16, function(ncl) table(cutree(data.clusts,  
ncl), data$Zamestnanci))  
  
names(cl_sizes) = 6:16  
  
cl_sizes
```

Výstup je pak k dispozici v příloze na straně číslo XI. Po jeho detailním prozkoumání lze konstatovat, že ani zde se nepodařilo nalézt jednoznačně prokazatelný vztah mezi

velikostí subjektu a výstupy ukazateli ekonomické výkonnosti podniku. Nejobsáhlejší shluk objektů nesoucí taxonomii 300 (100 až 1 000 zaměstnanců) se při zvyšování počtu shluků začne výrazněji rozdělovat mezi dva největší shluky. Do druhého shluku se zařadí celých 50 % subjektů z kategorie 400, což by mohlo vest k závěru, že se jedná o shluk větších společností, avšak z druhé strany jsou v tomto shluku významně zastoupeny také subjekty z kategorie 200 a všechny z kategorie mikropodniků do 10 zaměstnanců, což je v přímém protikladu s očekáváními.

3.2.3 Analýza výřezu dat

Jak již bylo zmíněno v předcházejících kapitolách, je žádoucí se samostatně věnovat ještě variantě, kdy ze vstupní datové matice odstraníme větší podíl odlehlých objektů a dostaneme se tak detailněji k analýze poměrně velkých shluků, které vznikly.

Ze vstupní datové matice eliminujeme dalších 12 objektů, jmenovitě: RWE Česká republika a.s., AGROTRADE, a.s., INEKON GROUP, a.s., Huawei Technologies (Czech) s.r.o., ČEZ, a. s., KASPER GROUP s.r.o., Holík International s.r.o., CIMEX INVEST s.r.o., ČESKÁ TELEVIZE, VÍTKOVICE HOLDING, a.s., Vysoká škola technická a ekonomická v Českých Budějovicích a InterCora, spol. s r.o. Odstraněním těchto subjektů získáme datovou matici čítající 78 objektů. Výsledné dendrogramy pro tyto výřezy dat jsou v příloze na straně číslo XII a v příloze na straně číslo XIII.

Z vizualizace na základě klasifikace ekonomické činnosti je patrné, že v dendrogramu nelze nalézt významnější oblasti, které by seskupovaly firmy pouze z jednoho oboru. Ve výšce okolo 10 nelze nalézt ani jeden shluk firem například čistě ze sekce zpracovatelského průmyslu, ač jde v datovém souboru o nejpočetněji zastoupenou skupinu, což poměrně jednoznačně svědčí v neprospěch existence vztahu mezi, v analytické části zvolenými, ukazateli ekonomické výkonnosti a oborem podnikání firmy.

V případě rozkladu výsledků klasifikace dle velikosti společnosti a pro rozmezí dvou až šestnácti shluků se nedaří vypořádat významnější trendy, neboť objekty se do jednotlivých skupin rozkládají takřka rovnoměrně v poměru ke svému zastoupení ve vstupním datovém souboru. Lze tedy i v tomto případě konstatovat, že ani ve výřezu

z původního datového souboru se nepodařilo nalézt jakoukoliv významnou vazbu a rozdělení subjektů do jednotlivých skupin probíhá rovnoměrně.

3.3 Ploché shlukování

Vedle hierarchických metod shlukování existují také ploché metody shlukování, které jsou graficky prezentovány pomocí množin v dvourozměrném prostoru, kdy se prvkům přiřazuje příslušnost do patřičné množiny.

Problematické plochého shlukování se věnuji okrajově, protože v námi zpracovávané úloze je proces interpretace dat a jejich výstupů komplikovanější a hodí se spíše pro testování příslušnosti subjektu k dané množině.

3.3.1 Stanovení počtu shluků

Pro všechny přístupy plochého shlukování je shodná nutnost stanovit cílový počet klastrů, skrze stanovení počtu medoidů (viz strana 31), kolem nichž bude docházet ke shlukování.

K tomuto účelu využijeme funkci *pam*, která je součástí knihovny *cluster* a provádí klasifikaci objektů kolem medoidů pomocí PAM algoritmu. Jako vstup funkce slouží již připravená matice vzdáleností a počet shluků, které budeme chtít vytvořit. Právě manipulací s tímto druhým argumentem můžeme testovat jednotlivé varianty řešení úlohy.

```
data.pam = pam(data.dist, 6)
```

Pomocí příkazu *table* můžeme porovnat výsledky řešení, která jsou výstupem funkce *hclust* a *pam*. Příkaz při třech shlucích pak bude vypadat následovně.

```
table(cutree(data.clusts, 6), data.pam$clustering)
```

Princip plochého shlukování je založen na shlukování objektů kolem medoidů, které fungují jako střed daného klastru a jedná se tudíž o typického zástupce daného shluku. Vzdálenost všech objektů v tomto shluku k medoidu je minimální. O které objekty konkrétně se jedná můžeme zjistit následujícím příkazem.

```
data$Firma[data.pam$id.med]
```

Výstupem jsou v našem případě společnosti KARSIT GROUP SE, G – TEAM a.s., CRYTUR, spol. s r.o., SYNOT W, a.s., LENOVO TECHNOLOGY B.V. a KAPRSCH

TELEMATIC SERVICES spol. s r.o. Při porovnání s dendrogramem výřezu dat je jasné patrné, že jeden z těchto shluků je jedno-objektový, tudíž odlehlý.

Avšak jedním z nejdůležitějších vizualizačních nástrojů funkce *pam* je graf zobrazující míru podobnosti (příslušnosti) objektů ve shluku, do nichž byly přiřazeny. Míra podobnosti je odvozena porovnáním podobnosti objektu s ostatními zástupci daného shluku. Vykreslený graf pro šest shluků je uveden v příloze na straně číslo XIV. Ideální jsou hodnoty blízké hodnotě jedna pro všechna pozorování (viz strana číslo 28). V našem případě jsou průměrné hodnoty míry příslušnosti objektu k danému shluku na úrovni 0.19, což indikuje, že v datech není dokonce žádná výraznější struktura a nemáme tudíž vhodně zvolený počet shluků.

Při dvou shlucích je průměrná míra příslušnosti na úrovni 0.55, což indikuje, že v datech byla nalezena smysluplná struktura. S přidáváním počtu clusterů se průměrná hodnota vazby dále snižuje, až při čtyřech shlucích klesne pod hranici 0.5. Hodnoty pod touto mezí již indikují, slabou, či umělou strukturu v datech.

V našem případě nenapomohlo k nalezení jasnější struktury v datech ani provedení dalšího výřezu ve vstupní datové matici. Byly eliminovány další pravděpodobně odlehlé objekty nacházejících se mimo nejrozsáhlejší v dendrogramech vizualizovaný shluk, avšak k výraznějšímu nárůstu indikace struktury v datech nedošlo.

Proto jako počet klastrů pro vizualizaci dat zvolíme tři, jde o množství shluků, kdy je v datech ještě patrná struktura.

3.3.2 Vizualizace dat

Data při plochem shlukování lze vizualizovat pomocí funkce *clusplot*, jenž je součástí knihovny *cluster*. Výstup níže uvedeného příkazu je k vidění v příloze na straně číslo XV.

```
clusplot(data.pam, "Cluster analysis", data.dist)
```

Graf nám jen dále vizuálně potvrdil informace, které jsme získali v předchozí kapitole při analýze struktury dat pro potřeby stanovení počtu shluků.

Je patrné, že máme vytvořen jeden klastř s poměrně silným jádrem, v němž se však také nacházejí dosti odlehlé hodnoty. Zbývající dva shluky jsou velmi roztažené a vazba mezi jednotlivými objekty tedy bude velmi slabá.

Při vytažení okrajů jednotlivých shluků dokonce vidíme, že dochází k výraznému průniku jednoho clusteru do druhého a dva body jsou dokonce natolik hraniční, že jsou v grafu vykresleny prakticky na sobě, a přesto mají rozdílnou příslušnost v rámci klastrů. Překryv může být do určité míry omezen eliminací silně odlehlého objektu LENOVO TECHNOLOGY B.V. (výrazně vpravo), nicméně pak jedna z hodnot přeteče přímo do jiného shluku.

Při užití plochého shlukování jsme došli k závěru, že strukturální shoda v použitých datech je nízká, což lze vizuálně dovodit také z grafu.

3.4 Interpretace výsledků

Nyní se soustředíme na výstupy návrhové části a agregaci dílčích závěrů jednotlivých užitých metod na námi zkoumaných datech.

V kapitole číslo 3.3.1, která se věnovala problematice stanovení počtu shluků pro potřeby plochého shlukování, jsme zjistili, že ve zdrojových datech lze jen velmi obtížně najít silnou strukturu, neboť průměrný koeficient míry příslušnosti objektu k danému klastru při počtu šesti shluků (výchozí hodnota pro klasifikaci dle v datovém souboru obsažených CZ-NACE skupin) dosahoval pouze hodnoty 0.19. Nejvyšších hodnot koeficientu jsme byli schopni logicky dosáhnout při shlukování do dvou klastrů. Na hraně smysluplnosti vychází ještě rozdělení do tří skupin. Je tak zřejmé, že nalezení přímé vazby mezi hodnotami ukazatelů ekonomické výkonnosti a oborem činnosti, či velikosti společnosti je v souboru použitých dat nerealizovatelné.

Nicméně graf plochého shlukování, jenž je uveden v příloze na straně číslo XV by nám mohl velmi dobře posloužit jako vodítko při rychlé identifikaci společností, jejichž ekonomické ukazatele se nějakým způsobem výrazněji odchyľují od běžných hodnot. Významná skupina společností se totiž shlukla poměrně velmi těsně, zatímco ostatní hodnoty jsou velmi výrazně vzdáleny a rozptýleny. Je však důležité si uvědomit, že tento graf neposkytuje vysvětlení, proč je konkrétní objekt odlehlý, interpretace takového výsledku by byla výsledkem hlubší analýzy.

Již v kapitolách věnujících se hierarchickému shlukování došlo na základě analýzy prezentovaných dendrogramů při interpretaci výsledků k vyvrácení obou zkoumaných hypotéz. Avšak nebyli jsme schopni v tu chvíli s jistotou rozhodnout o tom, jak silná je

v datech struktura. Po aplikaci metod plochého shlukování, můžeme i přes některé zajímavé vazby popsané například v kapitole 3.2.1 u zkoumaných dat obě v cílech vyřčené hypotézy považovat za nepotvrzené. Neopominutelným hlediskem také je, že hned v prvním kroku došlo k eliminaci výrazně odlehlých objektů. Provedení této eliminace je věnován jeden z úvodních odstavců kapitoly 3.2. Je důležité si uvědomit, že eliminace odlehlých objektů je běžnou praxí, která má za cíl zajistit, aby výstupy byly analyticky uchopitelné a nebyly zkresleny extrémními a statisticky nevýznamnými hodnotami. Na druhou stranu to však znamená, že samotné analýzy byly prováděny na výřezu dat. Ani to nevedlo k tomu, abychom získali konzistentní výsledky.

FINANČNÍ NÁROČNOST ŘEŠENÍ

Vyčíslení v této práci popsaného řešení vychází z jednorázových nákladů na vybudování a pořízení všech technologických prostředků umožňujících zpracování a procesování vstupní dat. Dále pak jde o variabilní náklady, které jsou dány administrativní činností spojenou s plněním datového skladu a zpracovávání vložených dat.

Mezi vynaloženými náklady je jediný nákup softwaru třetí strany, tímto nákupem je ABBYY FineReader, který je nezbytně nutný pro převod získaných účetních závěrek ve formě PDF dokumentů a rozličných PDF scanů do strojově čitelné podoby. Náklad na zakoupení časově neomezené licence vyhovující potřebám dané úlohy činí 5 399 Kč za Standard licenci [26], v případě požadavku na automatický převod dokumentů, který však pro tento typ úlohy nelze plně využít pak cena činí 8 099 Kč za Corporate licenci [26]. Pro potřeby této kalkulace budeme uvažovat Standard licenci, s jejíž pomocí byla tato práce prováděna.

Dalším jednorázovým investičním nákladem je naprogramování datového editoru. V tomto případě se datovým editorem rozumí v první řadě návrh architektury datového úložiště. Dále pak návrh a naprogramování aplikace zpracovávající digitalizovaná data a ukládající je do struktur datového úložiště. Datový editor má základní samoučící mechanismus, který při větším množství závěrek s různě užitou terminologií a strukturou výkazů výrazně urychluje administrativní práci, jenž vzniká při přidávání dalších subjektů. Tento mechanismus v dlouhodobém horizontu zefektivňuje proces přidávání nových dat a snižuje tak náklady spojené s administrativou. Pod datový editor pro potřeby této kalkulace zahrneme také analytický skript, který nad daty v uložišti provádí dávkově i jednotlivě výpočet poměrových ukazatelů uvedených v tabulce na straně číslo 48. Všechny výše popsané komponenty jsou vytvořeny pomocí programovacího jazyka PHP. Datové úložiště využívá relační databázi MySQL. Čas nutný k vytvoření výše uvedené aplikace zabral zhruba 8 MD. Při průměrné hodinové mzdě programátora v jazyce PHP, která činí 33 971 Kč hrubého měsíčně [27] můžeme tento náklad vyčíslit na 13 588 Kč. Při výpočtu uvažujme 20 pracovních dní v měsíci.

Další níže uvedené náklady jsou již variabilní a odvíjejí se zejména od kvality a dostupnosti podkladových dat. Ve své kalkulaci budu vycházet z osobní zkušeností

získaných průměrných hodnot, jenž by měly eliminovat extrémy a to oběma směry, neboť mezi nejkratším a nejdelším časem potřebným pro zpracování jednoho subjektu je v extrémním případě rozdíl až 30 minut.

V případě získávání potřebných údajů jsou nejčastějším informačním zdrojem Veřejný rejstřík a sbírky listin Ministerstva spravedlnosti a Český statistický úřad. Automatizovat dotazování na tyto instituce je problematické, zejména pak v případě Veřejného rejstříku, kdy účetní závěrky nejsou vždy identifikovány do té míry, aby jejich stahování mohl provádět automatický skript. Ne vždy se také jedná pouze o jediný soubor. Čas nutný k získání potřebných údajů zpravidla nepřesahuje 10 minut. Vyjma případů, kdy účetní závěrka není uveřejněna a je třeba se ji pokoušet dohledávat z dalších zdrojů. Čas potřebný pro získání dat o v této práci analyzovaných 100 subjektech tak vychází na cca 2 MD.

Digitalizace dat za využití ABBYY FineReaderu je nejchoulostivějším krokem celého procesu, protože právě zde se naplno projeví, v jaké kvalitě jsou zdrojová data k dispozici a případné chyby se zanáší i do dalších kroků. Průměrná doba potřebná k digitalizaci netextového dokumentu účetní závěrky (jde o naskanované dokumenty uložené jako PDF), kterých, jak je patrné z přílohy na straně číslo I, je naprostá většina zabere v průměru 15 minut. Tento krok sestává z lokalizace účetních výkazů ve výroční zprávě, kalibrace dokumentu a následné digitalizace do formátu tabulky. Abychom digitalizovali objem sta účetních závěrek dostáváme se na čas cca 3 MD.

Posledním manuálně prováděným krokem je nahrávání informací do datového úložiště prostřednictvím datového editoru. Náročnost tohoto kroku je podmíněna několika faktory. Největší vliv má struktura digitalizovaného souboru, kdy může v extrémních případech docházet k tomu, že výstupem digitalizace jsou složité tabulky, kde se buňky různě prolínají. Časovou náročnost zvyšuje také fakt, zda analyzovaná společnost již v databázi existuje, nebo je třeba její profil teprve vytvářet. Ovšem při objemu dat požadovaných pro zpracování této práce má tato proměnná pouze marginální vliv. Průměrný čas nutný k nahrání účetní závěrky pro nově zakládanou firmu s ohledem na postupné učení datového editoru je zhruba 10 minut. Nahrání sta firem do datového úložiště tak zabere cca 2 MD.

Všechny časy uváděné v oblasti administrativní práce s daty nezohledňují nutné přestávky a jsou chápány jako čistý čas strávený prací. Celkově se jedná o 7 MD nutných k získání, digitalizaci a nahrání účetních závěrek sta vybraných společností. Při průměrné mzdě ve výši 19 788 Kč v oblasti administrativy [28] lze nutné náklady vyčíslit na 6 926 Kč, pokud budeme uvažovat 20 pracovních dní v měsíci. Náklady na jednotku pak činí 70 Kč.

Vyčíslení náročnosti řešení

Položka	Částka
ABBYY FineReader	5 399 Kč
Datový editor	13 588 Kč
Zpracování dat (100 ks)	6 926 Kč
Součet	25 913 Kč

Tabulka 2 **Vyčíslení náročnosti řešení** (Zdroj: vlastní tvorba)

POTENCIÁL PRAKTICKÉHO UŽITÍ

Celá tato diplomová práce je psána zejména s cílem nalézt odpověď na dvě, v cílech vyřčené, hypotézy. Na první pohled tedy nemusí být jasně viditelné možnosti užití řešení v životním cyklu právnických osob. V této kapitole autor navrhuje možná řešení implementace analytické a návrhové části této práce, případně jejich dílčích celků.

V této práci představené řešení se zabývá klasifikací firem na základě ekonomické aktivity, proto je řešení téměř v celém objemu využitelné pro **benchmarking**. Za předpokladu, že by se podařilo vytvořit vstupní datové matice ze sledovaných konkurentů, nebo čistě oborové vstupní datové matice, tak podnik může na základě vybraných ekonomických ukazatelů pravidelně zjišťovat, zda při své současné ekonomické výkonnosti se stále nachází v požadovaných klastrech a nevzdaluje se ve sledovaných ukazatelích od požadovaných hodnot. Rozpětí sledovaných ukazatelů samozřejmě nutně nemusí být pojato tak široce, jako v této práci, ale společnost může a měla by sledovat pouze pro sebe podstatné hodnoty. Je však nutné zmínit, že benchmarking postavený výhradně na základě veřejných údajů z účetních závěrek sebou nese několik problémů. Nejzásadnějším je, že údaje jsou zpravidla neúplné nebo zkreslené nejrozličnějšími účetními praktikami. Při mezinárodním srovnání můžeme začít narážet na rozdílnost účetní legislativy jednotlivých zemí.

Druhou kapitolu této práce, tedy návrh procesu zpracování dat lze využít pro vybudování služby nabízející **analytické informace** o podnikatelských subjektech pohybujících se na trhu. V takovém případě by bylo nutné rozšířit záběr importovaných dat na kompletní účetní závěrky, neboť v rámci kroků činěných pro potřeby této práce jsem importoval pouze nezbytně nutná data. Toto rozšíření by samozřejmě navýšilo náklad na zpracovanou jednotku. Využitelná by tato služba mohla být pro nepřeborné množství analýz jednotlivých segmentů trhu, či při analýze obchodních partnerů, zejména v oblasti bonity. Byl by zde také potenciál v napojení na další veřejně dostupné rejstříky (například obchodní rejstřík, rejstřík živnostenského podnikání, insolvenční rejstřík, veřejný rejstřík dlužníků, spolehlivost plátce DPH, dotace z EU, veřejné zakázky atp.) a pokusit se tak nabídnout komplexní pohled na podnikatelskou aktivitu daného subjektu.

ZÁVĚR

Práce se věnovala potvrzení, či vyvrácení existence přímého kauzálního vztahu mezi ukazateli ekonomické výkonnosti firem a jejich klasifikací z pohledu provozované ekonomické činnosti a potvrzení, či vyvrácení existence přímého kauzálního vztahu mezi ukazateli ekonomické výkonnosti a velikosti firem dle prezentovaného počtu zaměstnanců.

V práci bylo konstatováno, že u zvolených dat nelze prokázat přímou spojitost mezi vybranými poměrovými ukazateli a klasifikací společností ze dvou výše zmíněných hledisek (viz strana 63). Výstupem testování dat v návrhové části jsme došli k závěru, že míra soudružnosti dat v rámci jednotlivých klastrů je nízká a hraniční z hlediska použití metod shlukové analýzy na takováto data. Ovšem při zúžení úlohy na tři clustery se dostáváme k míře koeficientu příslušnosti, kdy má smysl se zabývat interpretací výsledků (viz strana 61). Při výběru vstupních dat byla kladena snaha na zachování maximální objektivity. Proto byly subjekty zahrnuté do analýzy voleny na základě veřejné dostupného žebříčku českých 100 nejlepších firem (viz strana 33), jde o způsob, který byl zcela oddělen od autora práce, tudíž můžeme vzorek vstupních dat považovat za náhodně vybraný a reprezentativní.

K tomu abychom mohli konstatovat závěr ohledně zkoumaných hypotéz vedla celá řada kroků. V první řadě bylo nutné získat podkladová data. Při té příležitosti došlo ke zmapování informačních zdrojů, ze kterých lze potřebné informace čerpat a zejména k posouzení kvality firmami poskytovaných dat.

V dalším kroku pak docházelo k digitalizaci nashromážděných dat. Bylo nutné se seznámit s nabídkou na trhu dostupných programů s technologií OCR a z nich vybrat. Jako nejvhodnější nástroj byl zvolen ABBYY FineReader a následně byla kompletně převedena všechna nashromážděná data do strojově čitelného formátu (značkovací jazyk HTML).

Dále byla vytvořena aplikace, jejímž účelem bylo z převedených dat extrahovat informace nezbytné ke úspěšnému provedení finanční analýzy. Výstupy této finanční analýzy fungovaly jako vstupní datová matice (viz přílohy na stranách číslo IV, V a VI), na níž jsme ověřovali dvě vyřčené hypotézy. K finanční analýze byly využity v praxi

nejpoužívanější poměrové ukazatele ze čtyř základních oblastí (rentability, likvidity, aktivity a zadluženosti).

Posledním krokem bylo samotné ověřování hypotéz pomocí metod shlukové analýzy. Při tomto ověřování postupně došlo k ošetření prázdných hodnot, ke standardizaci dat, tak abychom porovnávali pouze srovnatelné proměnné. Takto předpřipravená data vstupovala do hierarchického aglomerativního shlukování. Při prvním shlukování nabýval výstupní dendrogram iracionální podoby, která prakticky znemožňovala jakékoliv okomentování získaných výsledků. Znamenalo to, že v datové matici se nachází některé velmi odlehlé hodnoty, jenž je nutné eliminovat – vyčistit data. Výstup shlukování vyčištěné vstupní datové matice již byl podroben analýze a okomentován. Jak bylo zmíněno v této kapitole, tak i přes některé zajímavé fenomény se nepovedlo prokázat žádný přímý kauzální vztah.

V jedné z posledních kapitol návrhové části byla data podrobena také metodám plochého shlukování. Výstupy této části jsou cenné zejména s ohledem na hledání ideálního počtu shluků, kdy bude míra příslušnosti objektu k danému shluku dosahovat maximálních hodnot. Bylo tak zjištěno, že vazba mezi objekty je na hranici smysluplnosti již při třech shlucích, tudíž je nesmyslné se data snažit testovat na hypotézami předpokládaný počet shluků.

V závěru práce krátce vyčísluji náklady na realizaci navržených postupů a nastiňuji možnosti praktického užití celého řešení, nebo jeho dílčích částí.

Na základě všeho výše uvedeného je zřejmé, že nebyla potvrzena ani jedna ze stanovených hypotéz. Tedy nebyla prokázána existence kauzálního vztahu mezi ukazateli výkonnosti firem a jejich klasifikací dle provozované ekonomické činnosti. Stejně tak nebyla potvrzena ani druhá hypotéza, a to, že existuje vazba mezi ukazateli ekonomické výkonnosti firem a jejich klasifikací z pohledu velikosti firmy.

SEZNAM POUŽITÝCH ZDROJŮ

- [1] KNÁPKOVÁ, Adriana, Drahomíra PAVELKOVÁ a Karel ŠTEKER. Finanční analýza: komplexní průvodce s příklady. 2., rozš. vyd. Praha: Grada, 2013. Prosperita firmy. ISBN 978-80-247-4456-8.
- [2] O obchodních společnostech a družstvech (zákon o obchodních korporacích). In: Sbírka zákonů. Sbírka zákonů České republiky, 2012, ročník 2012, částka 1370, číslo 90. Dostupné také z: portal.gov.cz/app/zakony/zakon.jsp?page=0&nr=90~2F2012&rpp=15
- [3] O účetnictví. In: Sbírka zákonů. Sbírka zákonů České republiky, 1991, ročník 1991, částka 2802, číslo 563. Dostupné také z: portal.gov.cz/app/zakony/zakon.jsp?page=0&nr=563~2F1991&rpp=15
- [4] KUČEROVÁ, Dagmar. Účetní závěrka se zasílá do obchodního rejstříku. In: Podnikatel.cz: Průvodce vaším podnikáním [online]. 2016 [cit. 2016-12-28]. Dostupné z: www.podnikatel.cz/clanky/ucetni-zaverka-se-zasila-do-obchodniho-rejstriku/
- [5] Technická specifikace pro předávání digitalizovaných listin do Sbírky listin veřejného rejstříku. Veřejný rejstřík a Sbírka listin [online]. Praha: Ministerstvo spravedlnosti České republiky, 2014 [cit. 2016-12-28]. Dostupné z: or.justice.cz/ias/ui/specifikaceSL
- [6] RŮČKOVÁ, P., D. PAVELKOVÁ a K. ŠTEKER. Finanční analýza: metody, ukazatele, využití v praxi. 4., aktualiz. vyd. Praha: Grada, 2011. 143 s. ISBN 978-80-247-3916-8
- [7] EIKVIL, Line. OCR - Optical Character Recognition [online]. [cit. 2017-01-14]. Dostupné z: www.nr.no/~eikvil/OCR.pdf
- [8] ABBYY FineReader 12 Professional. ABBYY [online]. 2016 [cit. 2017-01-09]. Dostupné z: www.abbyy.cz/products/personal/finereader/professional/
- [9] LACKO, Ľuboslav. *Databáze: datové sklady, OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle*. Brno, 2003. ISBN 80-722-6969-0.

- [10] ETL (Extract-Transform-Load). *Data Integration Info: Quick view on world of data* [online]. 2015 [cit. 2016-12-29]. Dostupné z: www.dataintegration.info/etl
- [11] HLAVENKA, Jiří. *Vytváříme WWW stránky*. 6. aktualiz. vyd. Praha: Computer Press, 2002. ISBN 80-722-6748-5.
- [12] JANOVSÝ, Dušan. *Jak psát web* [online]. 1999-2016 [cit. 2016-12-29]. ISSN 1801-0458. Dostupné z: www.jakpsatweb.cz
- [13] ZRALÝ, Jiří. JQuery pro začátečníky. In: *Digitální citron* [online]. 2008 [cit. 2016-12-29]. Dostupné z: citron.blueboard.cz/clanek/jquery-pro-zacatecniky-zaklady
- [14] LAHVIČKA, Jiří. PHP – základní informace. *Interval.cz* [online]. 2000 [cit. 2016-12-29]. Dostupné z: www.interval.cz/clanky/php-zakladni-informace/
- [15] ZAJÍC, Petr. MySQL (1) - pestrý svět databází. *Linuxsoft.cz* [online]. 2005 [cit. 2016-12-29]. Dostupné z: www.linuxsoft.cz/article.php?id_article=731
- [16] *Apache Friends* [online]. 2016 [cit. 2017-01-05]. Dostupné z: www.apachefriends.org/index.html
- [17] ŘEZANKOVÁ, Hana, Dušan HÚSEK a Václav SNÁŠEL. *Shluková analýza dat*. 2., rozš. vyd. Praha: Professional Publishing, 2009. ISBN 978-80-86946-81-8.
- [18] HEBÁK, Petr. *Vícerozměrné statistické metody*. Praha: Informatorium, 2005. ISBN 80-733-3039-3.
- [19] Klasifikace ekonomických činností (CZ-NACE). *BusinessInfo: Oficiální portál pro podnikání a export* [online]. 2017 [cit. 2017-01-09]. Dostupné z: www.businessinfo.cz/cs/clanky/klasifikace-ekonomickych-cinnost-cz-nace-3101.html
- [20] Vysledkyc100n2015. *COMENIUS* [online]. 2015 [cit. 2017-02-11]. Dostupné z: www.comenius.cz/vysledkyc100n2015
- [21] Klasifikace ekonomických činností (CZ-NACE). *Český statistický úřad: Statistický metainformační systém* [online]. Praha, 2017 [cit. 2017-02-11]. Dostupné z: apl.czso.cz/iSMS/klasstru.jsp?kodcis=80004

- [22] Třídění podniku. *Podnikátor* [online]. 2016 [cit. 2017-02-12]. Dostupné z: www.podnikator.cz/zacatek-podnikani/zalozeni-spolecnosti/n:16644/Trideni-podniku
- [23] Kategorie počtu zaměstnanců (metodika OECD). *Český statistický úřad: Statistický metainformační systém* [online]. Praha, 2017 [cit. 2017-02-12]. Dostupné z: apl.czso.cz/iSMS/cisdet.jsp?kodcis=579&strana=1
- [24] Rozdíly v účetní závěrce podnikatelských subjektů v České republice. *Účetní kavárna* [online]. Praha, 2014 [cit. 2017-04-18]. Dostupné z: www.ucetnikavarna.cz/archiv/dokument/doc-d43952v55168-rozdily-v-ucetni-zaverce-podnikatelskych-subjektu-v-ceske/
- [25] What is R? *The R Project for Statistical Computing* [online]. 2017 [cit. 2017-04-28]. Dostupné z: www.r-project.org/about.html
- [26] Pricing. *ABBYY* [online]. 2016 [cit. 2017-05-19]. Dostupné z: http://www.abbyy.cz/products/smb_enterprise/finereader14/pricing/
- [27] Programátor PHP. *Platy.cz* [online]. 2017 [cit. 2017-05-19]. Dostupné z: <http://www.platy.cz/platy/informacni-technologie/programator-php>
- [28] Administrativní pracovník, referent. *Platy.cz* [online]. 2017 [cit. 2017-05-19]. Dostupné z: <http://www.platy.cz/platy/administrativa/administrativni-pracovnik-referent>
- [29] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2017). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.6.

SEZNAM TABULEK

Tabulka 1 Vzorce pro výpočet finanční analýzy	48
Tabulka 2 Vyčíslení náročnosti řešení.....	67
Tabulka 3 Výstup analýzy dat, část 1	IV
Tabulka 4 Výstup analýzy dat, část 2	V
Tabulka 5 Výstup analýzy dat, část 3	VI

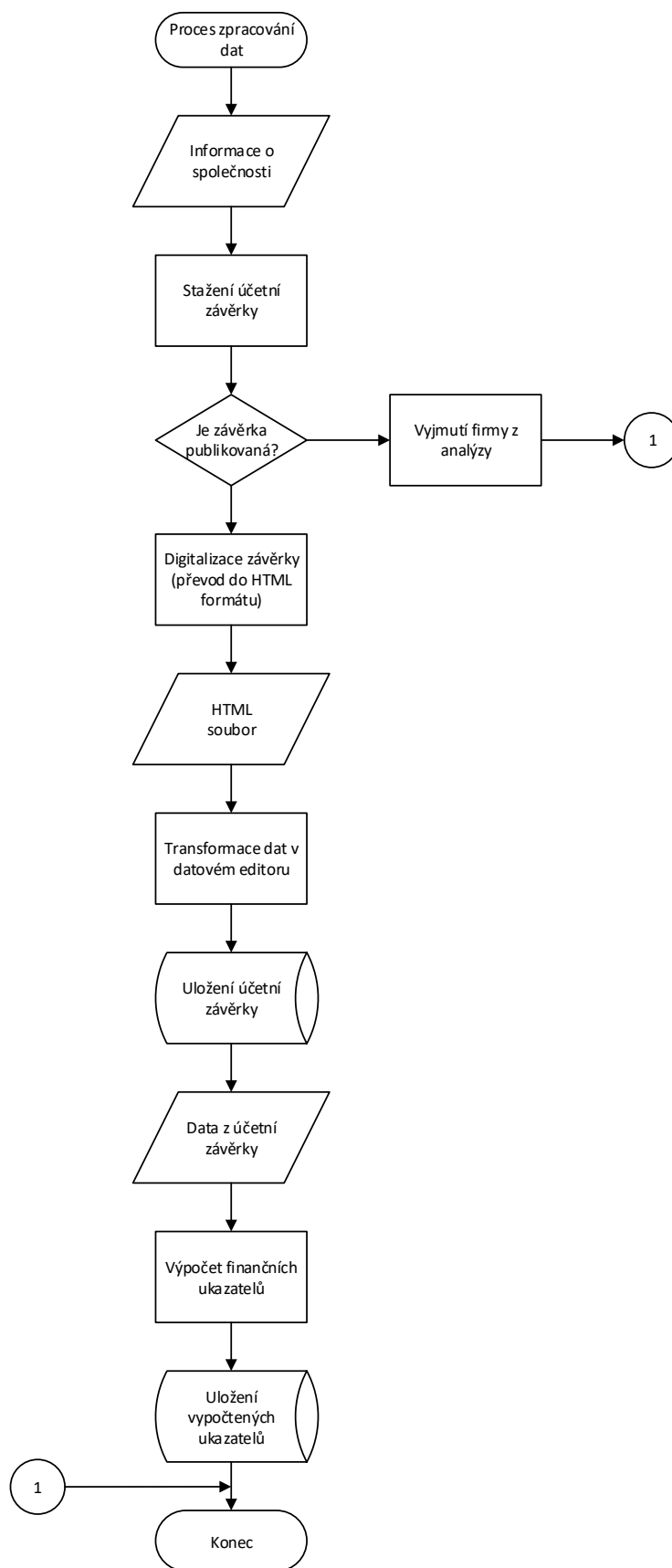
SEZNAM PŘÍLOH

Obrázek 1 Kvalita a dostupnost zdrojových dat	I
Obrázek 2 Vývojový diagram zpracování dat	II
Obrázek 3 Databázové schéma datového editoru	III
Tabulka 3 Výstup analýzy dat, část 1	IV
Tabulka 4 Výstup analýzy dat, část 2	V
Tabulka 5 Výstup analýzy dat, část 3	VI
Obrázek 4 Hierarchické shlukování – CZ-NACE	VII
Obrázek 5 Hierarchické shlukování – CZ-NACE	VIII
Obrázek 6 Hierarchické shlukování – dle velikosti	IX
Obrázek 7 Zastoupení činností v jednotlivých shlucích	X
Obrázek 8 Mediány ukazatelů v jednotlivých shlucích	X
Obrázek 9 Zastoupení subjektů dle velikosti při různém počtu shluků	XI
Obrázek 10 Dendrogram výřezu dat dle CZ-NACE	XII
Obrázek 11 Dendrogram výřezu dat dle velikosti	XIII
Obrázek 12 Graf míry příslušnosti objektu do clusteru	XIV
Obrázek 13 Ploché shlukování	XV

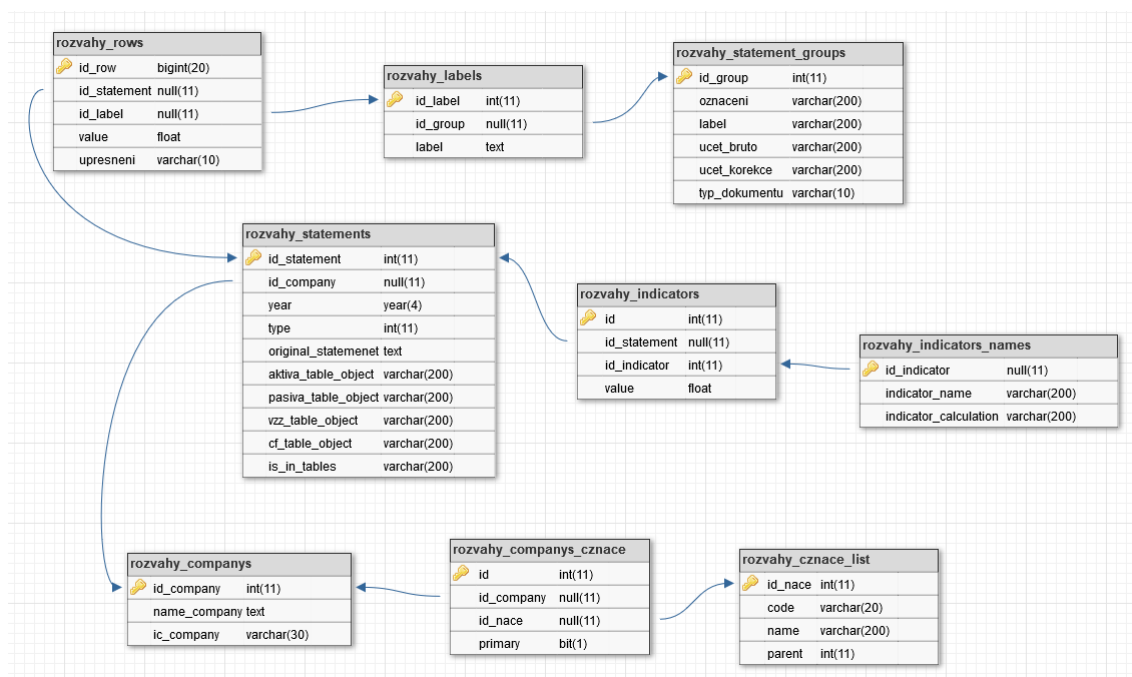
PŘÍLOHY



Obrázek 1 **Kvalita a dostupnost zdrojových dat** (Zdroj: vlastní tvorba)



Obrázek 2 Vývojový diagram zpracování dat (Zdroj: vlastní tvorba)



Obrázek 3 Databázové schéma datového editoru (Zdroj: vlastní tvorba)

Firma	CZN	Z	ROA	ROE	ROS	ROI	L3	L2	L1	OA	ODM	DOZ	DOP	DOZá	DI	ÚK
ŠKODA AUTO A.S.	C	510	17%	26%	10%	23%	1.43	1.12	0.94	1.55	2.93	17	14	89	0.72	1
AGROFERT, A.S.	G	320	7%	8%	13%	7%	1.15	0.98	0.10	0.48	0.58	17	92	109	0.22	1
ČEZ, A. S.	D	470	6%	13%	309%	7%	0.79	0.41	0.03	0.02	0.04	22	1645	0	1.35	1
BEST, A.S.	C	330	0%	0%	-1%	0%	2.57	0.61	0.06	0.53	0.65	58	23	54	0.41	0
MOUNTFIELD A.S.	G	420	2%	3%	0%	1%	1.28	0.33	0.03	1.45	2.99	78	30	117	5.19	0
KARLOVARSKÉ MINERÁLNÍ VODY, A. S.	C	330	7%	7%	6%	5%	1.91	1.12	0.08	0.68	0.89	52	68	85	0.70	1
KPMG ČESKÁ REPUBLIKA, S.R.O.	M	340	0%	41%	5%	41%	1.12	1.08	0.41	1.40	13.70	7	135	503	5.12	0
ČEPS, A.S.	D	340	6%	7%	10%	5%	1.59	1.53	1.09	0.50	0.55	2	21	64	0.45	1
FINCENTRUM A.S.	K	310	2%	2%	2%	1%	1.14	1.14	0.58	0.61	0.75	1	51	115	0.51	0
STUDENT AGENCY K.S.	H	420	14%	1%	0%	0%	0.84	0.30	0.08	1.72	3.18	2	25	153	2.42	1
LINET SPOL. S R.O.	C	340	6%	33%	5%	19%	10.02	0.69	0.07	0.81	1.37	34	105	210	6.44	1
VÍTKOVICE HOLDING, A.S.	C	210	1%	1%	86%	1%	3.76	2.49	0.53	0.01	0.01	241	375	679	0.01	1
SIKO KOUPELNY A.S.	G	340	5%	2%	0%	1%	5.93	2.53	0.67	2.10	33.85	87	49	29	2.41	0
KAREL HOLOUBEK - TRADE GROUP A.S.	G	310	2%	3%	2%	1%	1.28	0.95	0.10	0.47	0.69	1	113	157	2.19	0
S GROUP HOLDING, A.S.	M	230	0%	-6%	-5492%	-2%	1.54	1.39	0.28	0.00	0.00	39649	336587	134	2.84	1
AVE CZ ODPADOVÉ HOSPODÁŘSTVÍ S.R.O.	E	410	9%	15%	9%	9%	1.21	1.14	0.36	0.68	0.87	2	66	149	1.40	1
TOP HOTELS GROUP A.S.	G	310	0%	-6%	-11%	-3%	1.60	1.54	0.12	0.28	0.32	6	136	173	1.05	0
LINDE GAS A.S.	C	340	34%	34%	30%	33%	2.83	2.58	0.06	0.94	1.37	10	106	87	0.15	1
AAA AUTO INTERNATIONAL A.S.	G	420	7%	13%	4%	7%	1.82	0.30	0.20	1.58	2.19	48	4	0	1.12	0
RWE ČESKÁ REPUBLIKA A.S.	M	330	7%	8%	486%	8%	1.16	1.16	0.09	0.01	0.02	0	4242	5932	0.19	1
BOHEMIA SEKT, S.R.O.	C	330	23%	33%	18%	33%	2.67	1.73	0.22	0.99	4.62	91	161	164	0.81	1
BONECO A.S.	C	310	3%	3%	2%	3%	3.08	2.45	1.29	1.47	2.96	25	46	51	0.21	1
ŘÍZENÍ LETOVÉHO PROVOZU ČESKÉ REPUBLIKY, STÁTNÍ PODNIK	H	410	7%	6%	9%	6%	6.35	6.31	4.72	0.59	0.98	1	59	267	0.09	1
EXCON, A.S.	M	240	7%	11%	7%	10%	2.78	2.01	0.40	0.99	13.76	12	197	202	0.67	1
SPORTSIMO S.R.O.	G	430	6%	49%	3%	36%	1.57	0.18	0.04	1.49	17.58	191	19	163	9.39	1
OKIN GROUP, A.S.	M	430	2%	6%	1%	4%	0.99	0.83	0.26	1.93	4.86	1	65	320	2.49	0
CZ LOKO, A.S.	C	340	6%	9%	3%	8%	1.52	1.00	0.00	1.43	3.24	48	93	116	0.93	1
SCANSERVICE A.S.	S	240	32%	58%	8%	54%	3.54	3.30	1.37	3.30	17.33	5	41	38	0.70	1
LENOVO TECHNOLOGY B.V. ORGANIZAČNÍ SLOŽKA	M	230	18%	100%	4%	46%	0.76	0.76	0.01	3.71	142.06	0	42	90	6.48	1

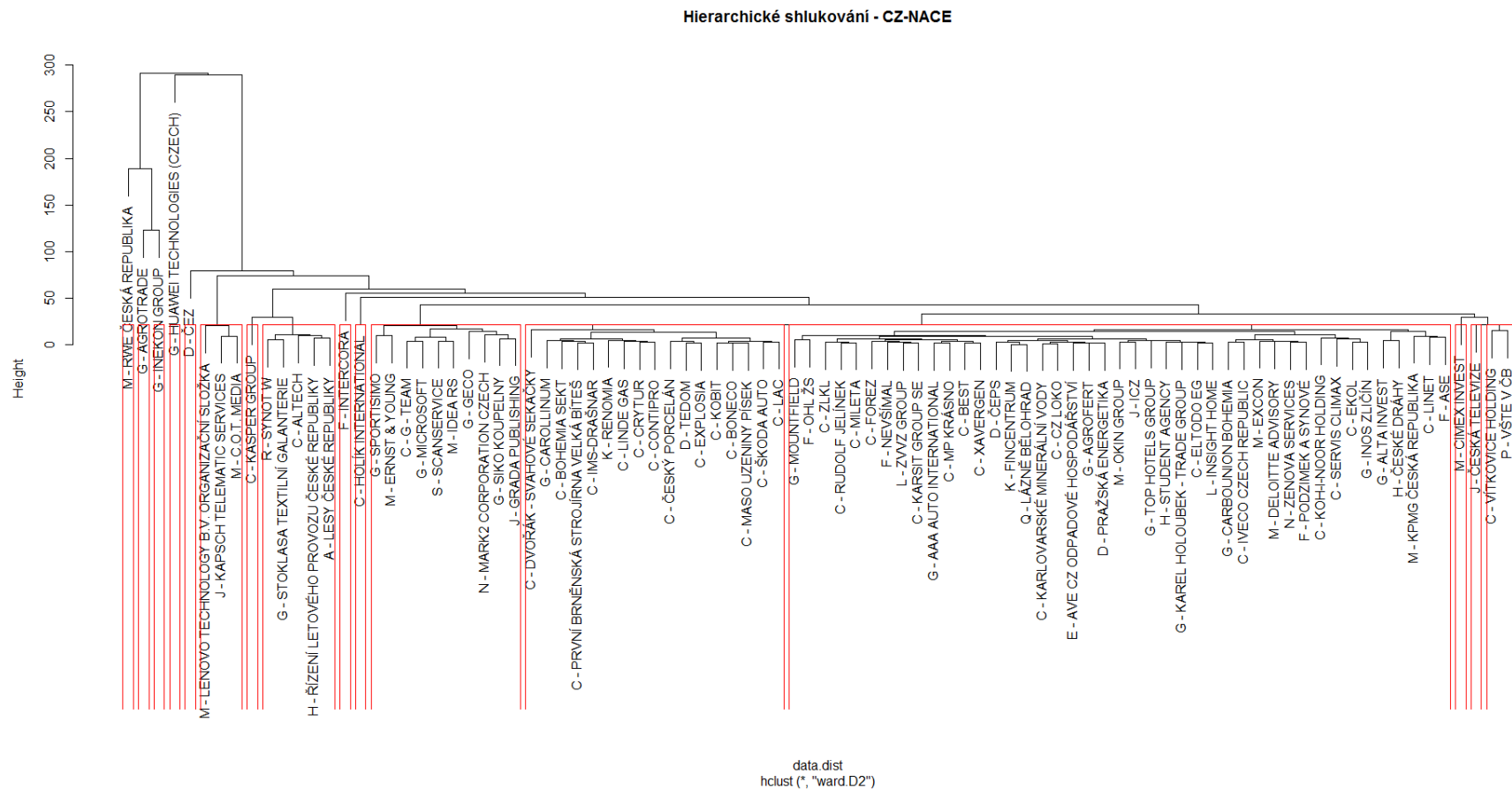
Tabulka 3 Výstup analýzy dat, část 1 (Zdroj: vlastní tvorba)

Firma	CZN	Z	ROA	ROE	ROS	ROI	L3	L2	L1	OA	ODM	DOZ	DOP	DOZá	DI	ÚK
AGROTRADE, A.S.	G	210	0%	20%	636%	14%	2.29	1.04	0.36	0.02	0.04	5796	3131	98	0.83	0
KAPSCH TELEMATIC SERVICES SPOL. S R.O.	J	310	50%	81%	19%	81%	2.07	1.95	0.32	2.12	81.56	6	129	104	1.00	1
ZENOVA SERVICES S.R.O.	N	310	2%	1%	0%	0%	1.82	1.76	0.00	2.49	10.59	2	91	115	1.64	0
LESY ČESKÉ REPUBLIKY, S.P.	A	450	9%	8%	46%	8%	7.14	7.00	5.87	0.17	0.19	5	39	102	0.03	1
GECO, A. S.	G	420	17%	58%	1%	58%	0.96	0.38	0.10	11.23	35.90	13	6	24	3.21	1
RENOMIA, A.S.	K	330	29%	44%	15%	42%	1.52	1.33	0.95	1.63	5.11	0	29	198	0.71	1
KOH-I-NOOR HOLDING A.S.	C	330	0%	14%	15%	14%	2.19	1.86	0.15	0.60	2.02	63	327	454	0.49	0
DELOITTE ADVISORY S.R.O.	M	340	4%	0%	0%	0%	1.29	1.17	0.04	2.69	19.19	9	100	132	0.00	1
ALTA INVEST, A.S.	G	120	2%	4%	2%	3%	1.30	0.20	0.19	0.46	2.70	73	3	583	2.59	1
ČESKÁ TELEVIZE	J	450	0%	0%	0%	0%	3.13	1.99	1.25	0.08	0.01	548	478	131	0.17	1
ZVVZ GROUP, A.S.	L	130	2%	2%	1%	2%	3.30	2.09	0.80	1.15	5.33	82	94	96	0.75	1
PRAŽSKÁ ENERGETIKA, A.S.	D	340	12%	20%	13%	15%	0.54	0.41	0.03	0.84	0.98	1	43	124	0.47	1
SYNOT W, A.S.	R	310	8%	6%	5%	2%	10.89	8.23	5.69	0.33	0.78	116	142	55	2.10	0
ENERGETICKÝ A PRŮMYSLVÝ HOLDING, A.S.	L	120	16%	27%	2499140%	17%	1.80	0.71	0.69	0.00	0.00	0	235525	16195	0.99	1
OHL ŽS, A.S.	F	410	-9%	-42%	-5%	-29%	1.04	0.80	0.42	1.73	6.75	1	57	161	4.17	0
ERNST & YOUNG, S.R.O.	M	340	14%	100%	4%	100%	1.36	1.02	0.15	2.43	27.46	26	80	200	12.51	1
G - TEAM A.S.	C	310	21%	64%	9%	64%	1.32	0.88	0.25	2.04	22.41	51	76	168	2.35	1
MICROSOFT S.R.O.	G	330	24%	85%	6%	85%	1.04	0.60	0.00	3.08	16.59	0	51	195	3.45	1
MARK2 CORPORATION CZECH A.S.	N	420	13%	42%	2%	29%	1.66	1.57	0.02	4.50	57.17	2	66	84	4.11	1
ASE, S.R.O.	F	240	17%	75%	7%	21%	2.79	2.38	0.23	1.96	5.86	7	93	60	4.20	1
CARBOUNION BOHEMIA, SPOL. S R.O.	G	230	14%	19%	4%	19%	2.08	2.00	0.23	2.75	10.87	3	80	47	0.55	1
XAVERGEN, A.S.	C	330	1%	1%	0%	1%	2.76	0.71	0.01	0.79	1.38	126	44	0	0.59	0
TEDOM, S.R.O.	D	340	4%	3%	4%	2%	5.18	3.06	1.38	0.48	0.83	121	101	87	0.74	0
INTERCORA, SPOL. S R.O.	F	210	5%	6%	17%	3%	0.40	0.20	0.00	0.12	0.14	25	177	2713	1.94	0
HUAWEI TECHNOLOGIES (CZECH) S.R.O.	G	330	4%	30%	1%	30%	1.34	0.92	0.21	3.18	697.78	32	60	99	10.24	1
MP KRÁSNO, A.S.	C	340	3%	5%	1%	4%	1.42	0.95	0.18	2.42	3.74	15	28	41	1.07	1
ICZ A.S.	J	330	8%	22%	5%	22%	1.30	1.23	0.50	1.46	6.80	6	105	211	1.72	1
ELTODO EG, A.S.	C	340	2%	1%	0%	1%	1.56	0.21	0.21	1.03	2.71	28	0	205	1.49	0
LÁZNĚ BĚLOHRAD A.S.	Q	330	2%	1%	1%	1%	1.29	1.16	0.45	0.53	0.59	5	34	125	0.65	0
MASO UZENINY PÍSEK, A.S.	C	330	9%	10%	4%	10%	2.77	2.24	1.27	2.16	4.25	15	29	35	0.24	1
RUDOLF JELÍNEK A.S.	C	310	4%	6%	5%	5%	2.29	1.20	0.14	0.66	1.58	147	146	179	0.85	1

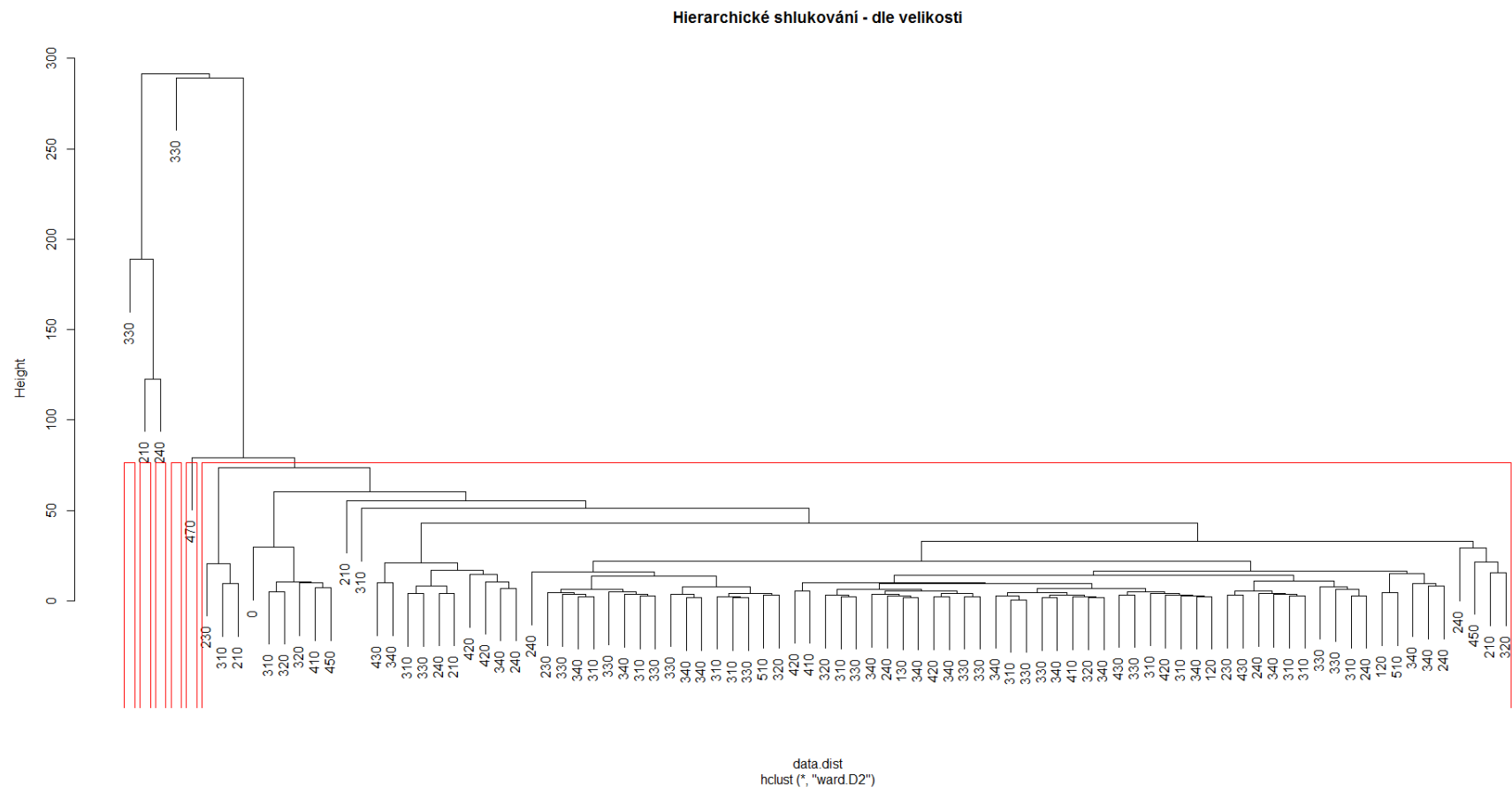
Tabulka 4 Výstup analýzy dat, část 2 (Zdroj: vlastní tvorba)

Firma	CZN	Z	ROA	ROE	ROS	ROI	L3	L2	L1	OA	ODM	DOZ	DOP	DOZá	DI	ÚK
PRVNÍ BRNĚNSKÁ STROJÍRNA VELKÁ BÍTEŠ, A.S.	C	340	18%	23%	13%	21%	2.89	1.22	0.34	1.14	2.59	101	53	104	0.55	1
IVECO CZECH REPUBLIC, A.S.	C	430	17%	22%	10%	22%	2.47	2.19	0.07	1.46	7.80	23	172	104	0.56	1
KARSIT GROUP SE	C	340	4%	8%	3%	6%	2.47	2.01	0.62	1.16	3.21	36	111	93	0.99	1
SERVIS CLIMAX A.S.	C	330	11%	14%	5%	12%	2.73	7.30	0.65	1.77	4.46	59	287	58	0.56	1
STOKLASA TEXTILNÍ GALANTERIE S.R.O.	G	320	1%	-1%	-2%	-1%	13.02	5.86	4.95	0.35	0.44	123	16	29	0.02	1
KASPER GROUP S.R.O.	C	0	8%	97%	0%	97%	8.36	8.36	8.36	0.00	0.00	0	0	990	11.52	1
GRADA PUBLISHING, A.S.	J	240	31%	31%	17%	31%	6.20	2.48	0.53	1.43	42.40	128	76	71	0.25	1
ČESKÉ DRÁHY, A.S.	H	510	-1%	-6%	-10%	-4%	0.38	0.22	0.12	0.28	0.31	17	30	635	0.00	0
ZKL, S.R.O.	C	320	8%	22%	6%	14%	2.77	1.40	0.00	1.01	2.47	103	106	98	2.53	1
CRYTUR, SPOL. S R.O.	C	310	26%	26%	18%	26%	4.70	2.56	0.56	1.17	2.17	63	59	62	0.21	1
CONTIPRO A.S.	C	330	25%	27%	23%	26%	2.66	2.21	0.85	0.95	1.59	25	77	120	0.22	1
IMS-DRAŠNAR S.R.O.	C	310	21%	30%	13%	21%	4.65	2.04	0.73	1.29	4.56	111	56	60	0.83	1
DVOŘÁK - SVAHOVÉ SEKAČKY S.R.O.	C	240	49%	59%	17%	46%	12.06	8.14	2.31	2.30	4.44	24	36	10	0.49	1
ALTECH, SPOL. S R.O.	C	320	19%	17%	17%	17%	9.19	8.23	7.92	0.89	3.62	31	10	67	0.10	1
MILETA A.S.	C	330	10%	15%	8%	10%	3.70	1.58	0.12	0.98	3.10	142	99	106	1.03	1
CIMEX INVEST S.R.O.	M	240	-7%	-120%	-52%	-19%	0.32	0.32	0.13	0.22	0.23	0	72	937	9.61	0
HOLÍK INTERNATIONAL S.R.O.	C	310	-60%	-289%	-57%	-263%	1.41	0.45	-0.28	1.67	3.99	82	63	123	1.99	0
PODZIMEK A SYNOVÉ S.R.O.	F	310	2%	5%	1%	4%	2.19	2.11	0.21	1.57	16.26	7	173	112	1.95	1
CAROLLINUM S.R.O.	G	230	30%	40%	13%	39%	2.99	0.65	0.43	1.80	11.46	128	12	75	0.69	1
EKOL, SPOL. S R.O.	C	310	-3%	4%	3%	4%	2.48	2.14	0.51	0.74	4.16	48	262	200	0.68	1
HIT OFFICE GROUP A.S.	G	0	0%	0%	0%	0%	1516.00	1516.00	516.00	0.00	0.00	0	0	54	0.00	1
VŠTE České Budějovice	P	320	1%	1%	23%	1%	4.35	4.20	3.58	0.04	0.05	54	389	1034	0.09	1
INEKON GROUP, A.S.	G	240	7%	59%	98%	55%	0.95	0.17	0.03	0.07	0.29	3244	649	706	8.13	1
IDEA RS S.R.O.	M	210	38%	63%	12%	60%	2.06	2.06	1.02	2.48	27.97	0	66	82	0.96	1
EXPLOSIA A.S.	C	340	4%	5%	7%	5%	4.13	2.43	1.29	0.61	0.98	91	61	92	0.27	1
NEVŠÍMAL, A. S.	F	240	2%	4%	1%	4%	2.50	1.99	0.31	1.50	4.93	33	110	75	2.44	0
ČESKÝ PORCELÁN, A.S.	C	330	0%	-1%	-1%	-1%	5.83	3.27	2.19	0.89	1.73	84	36	71	0.10	0
INOS ZLIČÍN, A.S.	G	240	5%	6%	6%	5%	2.86	2.49	1.08	0.56	2.34	61	235	98	0.51	1
C.O.T. MEDIA S.R.O.	M	210	18%	27%	4%	25%	1.93	1.93	0.64	3.61	96.07	0	60	67	1.04	1
LAC, S.R.O.	C	320	19%	20%	9%	19%	3.74	2.27	1.51	1.71	4.76	51	26	61	0.28	1
KOBIT, SPOL. S R.O.	C	310	9%	12%	5%	12%	2.50	2.27	1.34	1.60	8.80	16	68	85	0.55	1
INSIGHT HOME, A.S.	L	120	3%	10%	0%	9%	1.46	1.02	0.14	0.00	0.00	0	0	168	2.21	1
FOREZ S.R.O.	C	340	9%	10%	6%	7%	4.95	2.18	0.11	1.19	2.90	71	75	60	0.55	1

Tabulka 5 Výstup analýzy dat, část 3 (Zdroj: vlastní tvorba)



Obrázek 4 Hierarchické shlukování – CZ-NACE (Zdroj: vlastní tvorba)



Obrázek 6 **Hierarchické shlukování – dle velikosti** (Zdroj: vlastní tvorba)

clusters_size	A	C	D	E	F	G	H	J	K	L	M	N	P	Q	R	S
1	0	14	1	0	0	1	0	0	1	0	0	0	0	0	0	0
2	0	16	2	1	4	8	2	1	1	2	4	1	0	1	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
5	0	1	0	0	0	4	0	1	0	0	2	1	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
7	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0
8	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0
9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
11	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
15	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

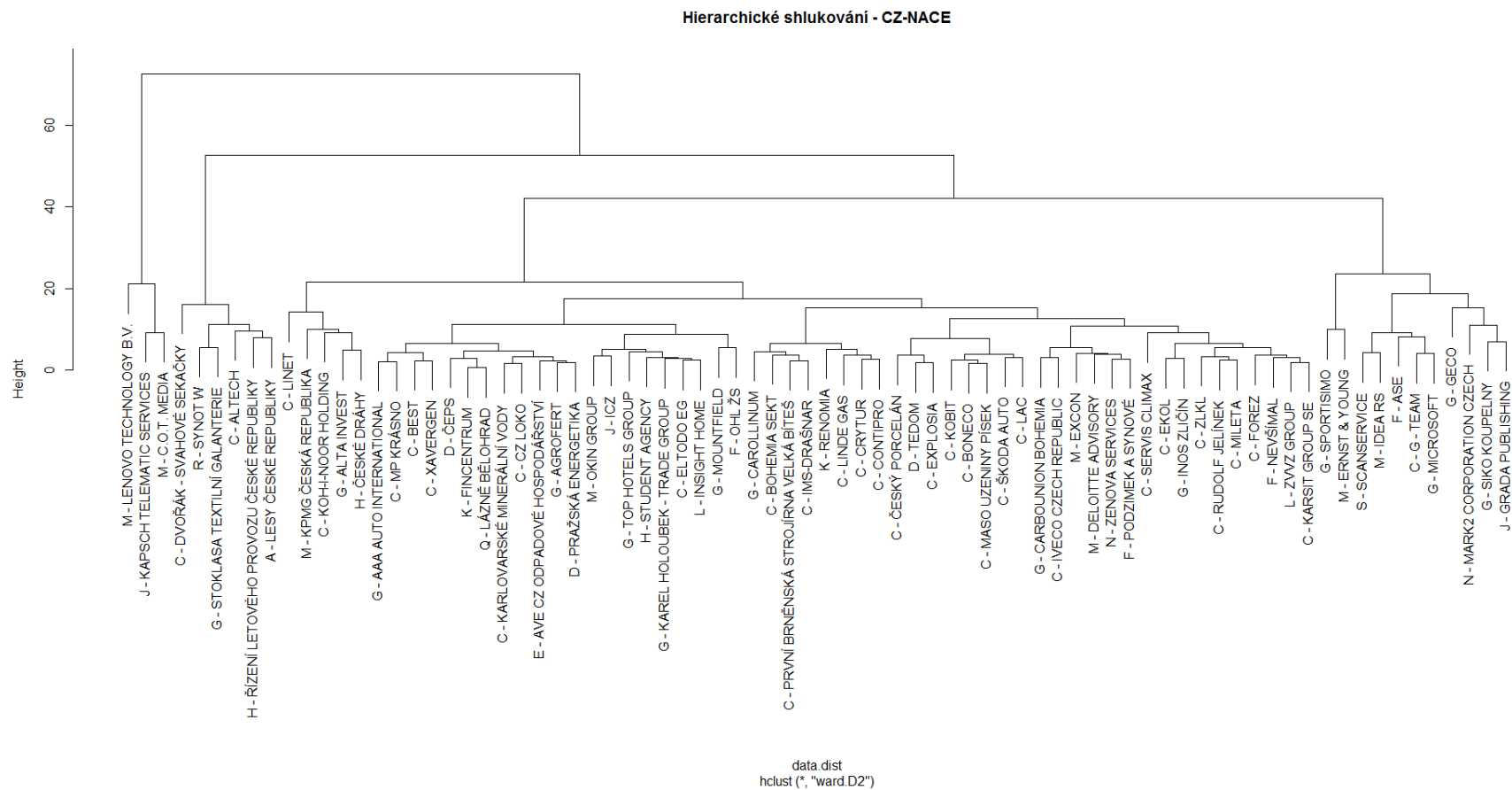
Obrázek 7 Zastoupení činností v jednotlivých slucích (Zdroj: vlastní tvorba)

	Group.1	ROA	ROE	ROS	ROI	Likvidita..III..	Likvidita..II..	Likvidita..I..	Obrat.aktiv
1	1	0.188727000	0.2600970	0.12661600	0.21312100	2.989240	2.268320	0.94705600	1.29116000
2	2	0.038459300	0.0604345	0.03294920	0.05274910	1.816520	1.160740	0.18267600	1.01439000
3	3	0.059099500	0.1274660	3.09364000	0.06853370	0.792545	0.411507	0.02740790	0.01753150
4	4	0.009740545	0.0084316	0.54497950	0.00842881	4.053500	3.345320	2.05030450	0.02327358
5	5	0.188156500	0.5790905	0.04918620	0.55670650	1.615395	1.293140	0.20010250	2.45563500
6	6	0.072284800	0.0846048	4.86415000	0.08449830	1.161030	1.158540	0.09228900	0.01456160
7	7	0.075897000	0.0582452	0.09057360	0.05703890	9.188620	7.001320	5.68684000	0.34670800
8	8	0.181739000	0.8113890	0.03643080	0.45758000	1.927620	1.927620	0.31977800	3.60521000
9	9	0.004724500	0.2048910	6.36349000	0.14434300	2.290050	1.037790	0.36122300	0.01693270
10	10	0.000000000	0.0000000	0.00000000	0.00000000	3.126150	1.993920	1.25331000	0.07830240
11	11	0.053119700	0.0632101	0.17275900	0.03375100	0.400670	0.195001	0.00313523	0.12398200
12	12	0.038353700	0.3043540	0.00827749	0.30435400	1.336470	0.919951	0.21394500	3.18086000
13	13	0.077369200	0.9685930	0.00000000	0.96859300	8.363640	8.363640	8.36364000	0.00000000
14	14	-0.071011000	-1.1989100	-0.51851700	-0.18516700	0.319106	0.317570	0.13157700	0.21717700
15	15	-0.604168000	-2.8928000	-0.57282700	-2.63137000	1.410110	0.447899	-0.28453800	1.67258000
16	16	0.067393000	0.5905300	0.97857000	0.55321200	0.946956	0.172357	0.02706270	0.06596820

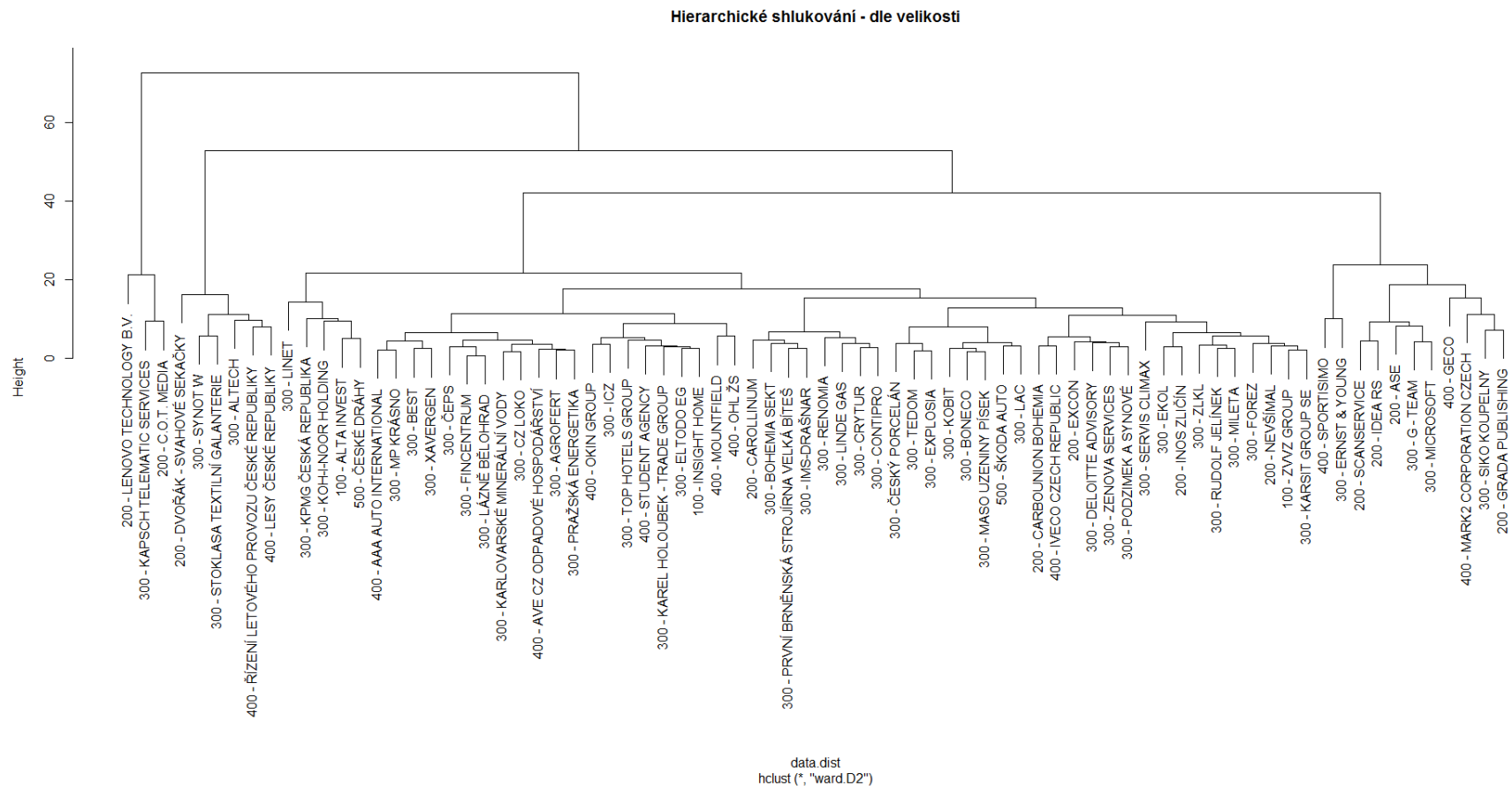
Obrázek 8 Mediány ukazatelů v jednotlivých slucích (Zdroj: vlastní tvorba)

\$`6`						
	0	100	200	300	400	500
1	1	3	15	51	13	2
2	0	0	0	0	1	0
3	0	0	0	1	0	0
4	0	0	1	0	0	0
5	0	0	0	1	0	0
6	0	0	1	0	0	0
\$`7`						
	0	100	200	300	400	500
1	1	3	13	50	13	2
2	0	0	0	0	1	0
3	0	0	0	1	0	0
4	0	0	2	1	0	0
5	0	0	1	0	0	0
6	0	0	0	1	0	0
7	0	0	1	0	0	0
\$`8`						
	0	100	200	300	400	500
1	0	3	13	47	11	2
2	0	0	0	0	1	0
3	0	0	0	1	0	0
4	1	0	0	3	2	0
5	0	0	2	1	0	0
6	0	0	1	0	0	0
7	0	0	0	1	0	0
8	0	0	1	0	0	0
\$`9`						
	0	100	200	300	400	500
1	0	3	12	47	11	2
2	0	0	0	0	1	0
3	0	0	0	1	0	0
4	1	0	0	3	2	0
5	0	0	2	1	0	0
6	0	0	1	0	0	0
7	0	0	1	0	0	0
8	0	0	0	1	0	0
9	0	0	1	0	0	0
\$`10`						
	0	100	200	300	400	500
1	0	3	12	46	11	2
2	0	0	0	0	1	0
3	0	0	0	1	0	0
4	1	0	0	3	2	0
5	0	0	2	1	0	0
6	0	0	1	0	0	0
7	0	0	1	0	0	0
8	0	0	0	1	0	0
9	0	0	0	1	0	0
10	0	0	1	0	0	0
\$`11`						
	0	100	200	300	400	500
1	0	3	9	42	8	2
2	0	0	0	0	1	0
3	0	0	3	4	3	0
4	0	0	0	1	0	0
5	1	0	0	3	2	0
6	0	0	2	1	0	0
7	0	0	1	0	0	0
8	0	0	1	0	0	0
9	0	0	0	1	0	0
10	0	0	0	1	0	0
11	0	0	1	0	0	0
\$`12`						
	0	100	200	300	400	500
1	0	3	7	41	7	2
2	0	0	0	0	1	0
3	0	0	2	1	1	0
4	0	0	3	4	3	0
5	0	0	0	1	0	0
6	1	0	0	3	2	0
7	0	0	2	1	0	0
8	0	0	1	0	0	0
9	0	0	1	0	0	0
10	0	0	0	1	0	0
11	0	0	0	1	0	0
12	0	0	1	0	0	0
\$`13`						
	0	100	200	300	400	500
1	0	3	7	41	7	2
2	0	0	0	0	1	0
3	0	0	2	1	1	0
4	0	0	3	4	3	0
5	0	0	0	1	0	0
6	0	0	0	3	2	0
7	0	0	2	1	0	0
8	0	0	1	0	0	0
9	0	0	1	0	0	0
10	0	0	0	1	0	0
11	1	0	0	0	0	0
12	0	0	0	1	0	0
13	0	0	1	0	0	0
\$`14`						
	0	100	200	300	400	500
1	0	3	7	41	7	2
2	0	0	0	0	1	0
3	0	0	1	1	1	0
4	0	0	3	4	3	0
5	0	0	0	1	0	0
6	0	0	0	3	2	0
7	0	0	2	1	0	0
8	0	0	1	0	0	0
9	0	0	1	0	0	0
10	0	0	0	1	0	0
11	1	0	0	0	0	0
12	0	0	1	0	0	0
13	0	0	0	1	0	0
14	0	0	1	0	0	0
\$`15`						
	0	100	200	300	400	500
1	0	0	2	14	0	1
2	0	3	5	27	7	1
3	0	0	0	0	1	0
4	0	0	1	1	1	0
5	0	0	3	4	3	0
6	0	0	0	1	0	0
7	0	0	0	3	2	0
8	0	0	2	1	0	0
9	0	0	1	0	0	0
10	0	0	1	0	0	0
11	0	0	0	1	0	0
12	1	0	0	0	0	0
13	0	0	1	0	0	0
14	0	0	0	1	0	0
15	0	0	1	0	0	0
\$`16`						
	0	100	200	300	400	500
1	0	0	2	14	0	1
2	0	3	5	27	7	1
3	0	0	0	0	1	0
4	0	0	1	1	0	0
5	0	0	3	4	3	0
6	0	0	0	1	0	0
7	0	0	0	3	2	0
8	0	0	2	1	0	0
9	0	0	1	0	0	0
10	0	0	0	0	1	0
11	0	0	1	0	0	0
12	0	0	0	1	0	0
13	1	0	0	0	0	0
14	0	0	1	0	0	0
15	0	0	0	1	0	0
16	0	0	1	0	0	0

Obrázek 9 Zastoupení subjektů dle velikosti při různém počtu shluků (Zdroj: vlastní tvorba)



Obrázek 10 Dendrogram výřezu dat dle CZ-NACE (Zdroj: vlastní zpracování)

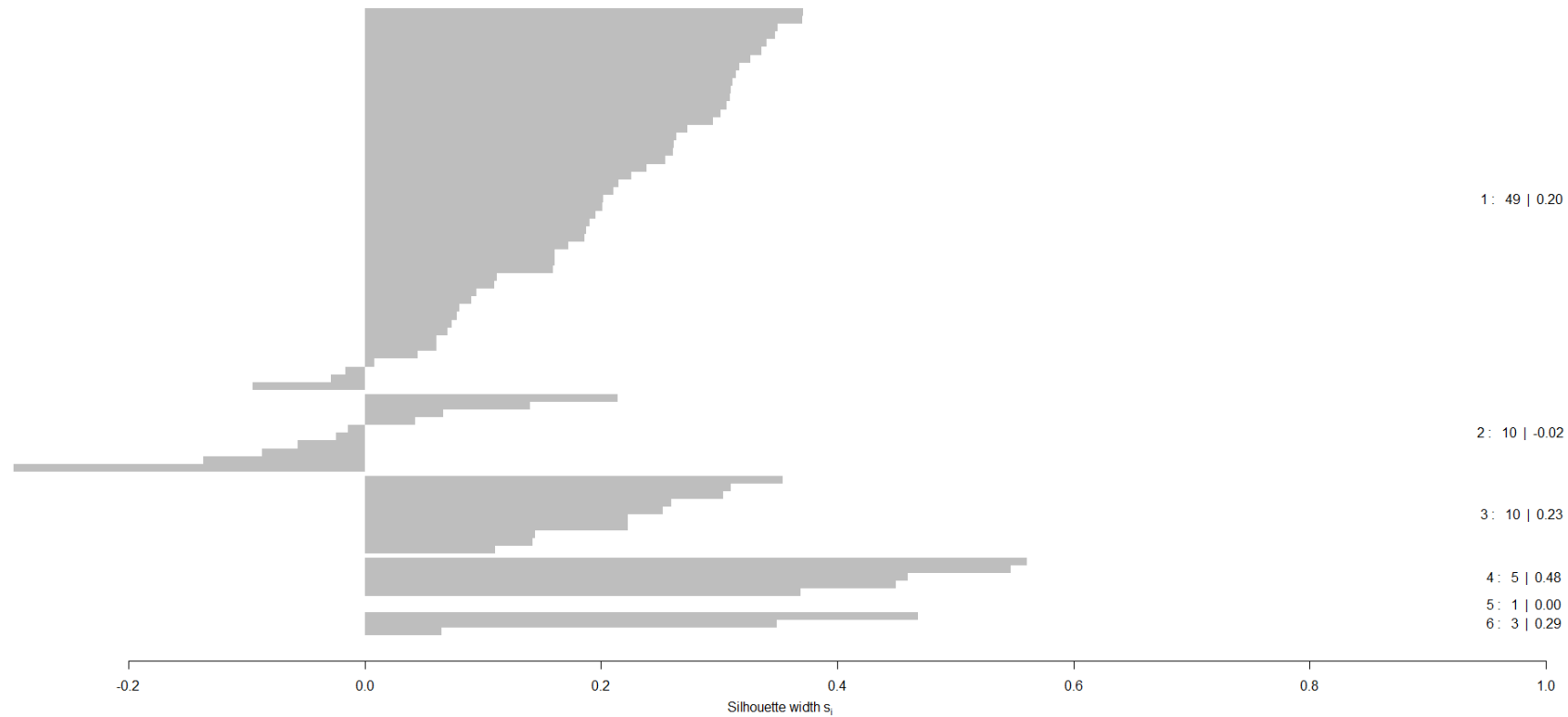


Obrázek 11 **Dendrogram výřezu dat dle velikosti** (Zdroj: vlastní zpracování)

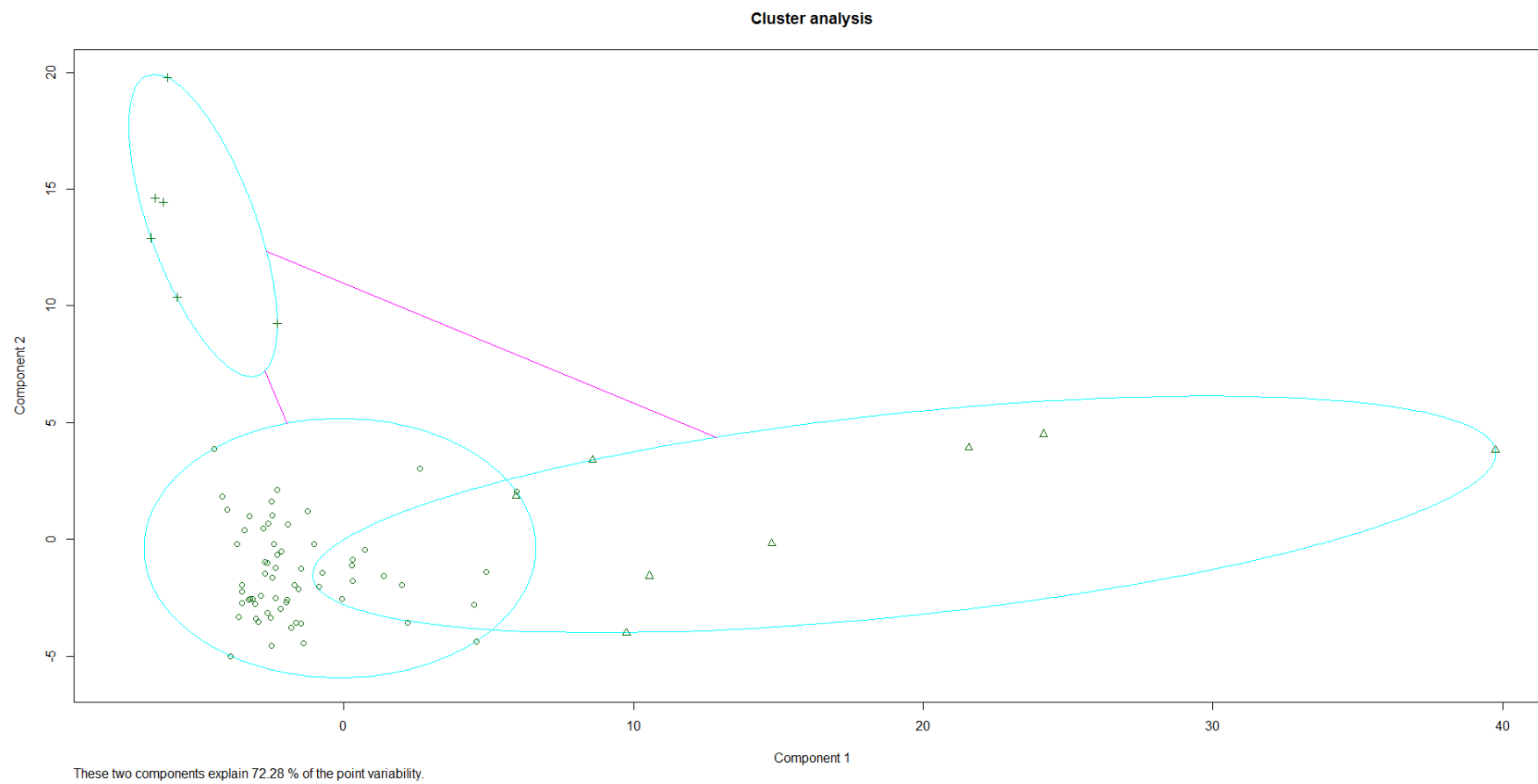
Silhouette plot of pam(x = data.dist, k = 6)

n = 78

6 clusters C_j
j : n_j | ave_{ie}



Obrázek 12 Graf míry příslušnosti objektu do clusteru (Zdroj: vlastní tvorba)



Obrázek 13 **Ploché shlukování** (Zdroj: vlastní tvorba)